


```

QY 61 ACTTCCTGTTTCGATTAGTTGGCAATTTAGTGGCAATTCAGTACAGACCCCGCAGATTCT 120
DB 61 ACTTCCTGTTTCGATTAGTTGGCAATTTAGTGGCAATTCAGTACAGACCCCGCAGATTCT 120
QY 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
DB 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
QY 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAAAATGCATGAGTCTCTCTGAATATGA 240
DB 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAAAATGCATGAGTCTCTCTGAATATGA 240
QY 241 CATCAACTGTGATATGATATCAATCTGCGGGTAAAGAGAGGTTCTTGTAGCGCG 300
DB 241 CATCAACTGTGATATGATATCAATCTGCGGGTAAAGAGAGGTTCTTGTAGCGCG 300
QY 301 TGAAGTGGCGGAGAGCGGAAATGTTGCGTGAATGTTCTGACTGCGCGTGAACAAGAGT 360
DB 301 TGAAGTGGCGGAGAGCGGAAATGTTGCGTGAATGTTCTGACTGCGCGTGAACAAGAGT 360
QY 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATACATCAACAAACGGTCAACC 420
DB 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATACATCAACAAACGGTCAACC 420
QY 421 GCGTGAACGATGATTCGTGACAGCACTTACTGCTCCGTAACATGATCTGGGTACTGT 480
DB 421 GCGTGAACGATGATTCGTGACAGCACTTACTGCTCCGTAACATGATCTGGGTACTGT 480
QY 481 CAGCGAAGACGTGCTAGCGTTGAAAGCTAACAGTTCAAGTTGGGAACTGAGCATCA 540
DB 481 CAGCGAAGACGTGCTAGCGTTGAAAGCTAACAGTTCAAGTTGGGAACTGAGCATCA 540
QY 541 CAGCGGTTGCTGTAATCGGCTCTGATGGGAGCAGTACAGCTGCGCGCAGCGAGTTCCG 600
DB 541 CAGCGGTTGCTGTAATCGGCTCTGATGGGAGCAGTACAGCTGCGCGCAGCGAGTTCCG 600
QY 601 CGCATGCTTCACTTCTGTGTAACCCAGGCAAAATTCATCTCCGCTGTAATCTGCTGAA 660
DB 601 CGCATGCTTCACTTCTGTGTAACCCAGGCAAAATTCATCTCCGCTGTAATCTGCTGAA 660
QY 661 GAAAATGACCGGCGCTGAGCTGAAACCGACGACCGTACTGTGAGATGATCCGCG 720
DB 661 GAAAATGACCGGCGCTGAGCTGAAACCGACGACCGTACTGTGAGATGATCCGCG 720
QY 721 TATTCGTAACATTTGGAATCTACGCGGATAGCGGAAATCATGCGCAATTCACGG 780
DB 721 TATTCGTAACATTTGGAATCTACGCGGATAGCGGAAATCATGCGCAATTCACGG 780
QY 781 TGAAGGTATCGCTTCTGCGGTGATCTGGAAGATTAATCGGCTTACCAACGCTCAAAA 840
DB 781 TGAAGGTATCGCTTCTGCGGTGATCTGGAAGATTAATCGGCTTACCAACGCTCAAAA 840
QY 841 AACGGCGCTTTTGAAGCGGTTTATTTTCAACCTTATTTCCAGATACGTAATCTATC 900
DB 841 AACGGCGCTTTTGAAGCGGTTTATTTTCAACCTTATTTCCAGATACGTAATCTATC 900
QY 901 GTTCGTTGTAACCTTCTTACTGCTTT 927
DB 901 GTTCGTTGTAACCTTCTTACTGCTTT 927

```

RESULT 2
U00096_46/c
WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096
Fragment Name Begin End
U00096_00 1 110000
U00096_01 100001 210000
U00096_02 200001 310000
U00096_03 300001 410000
U00096_04 400001 510000
U00096_05 500001 610000

```

QY U00096_06 600001 710000
DB U00096_07 700001 810000
QY U00096_08 800001 910000
DB U00096_09 900001 1010000
QY U00096_10 1000001 1110000
DB U00096_11 1100001 1210000
QY U00096_12 1200001 1310000
DB U00096_13 1300001 1410000
QY U00096_14 1400001 1510000
DB U00096_15 1500001 1610000
QY U00096_16 1600001 1710000
DB U00096_17 1700001 1810000
QY U00096_18 1800001 1910000
DB U00096_19 1900001 2010000
QY U00096_20 2000001 2110000
DB U00096_21 2100001 2210000
QY U00096_22 2200001 2310000
DB U00096_23 2300001 2410000
QY U00096_24 2400001 2510000
DB U00096_25 2500001 2610000
QY U00096_26 2600001 2710000
DB U00096_27 2700001 2810000
QY U00096_28 2800001 2910000
DB U00096_29 2900001 3010000
QY U00096_30 3000001 3110000
DB U00096_31 3100001 3210000
QY U00096_32 3200001 3310000
DB U00096_33 3300001 3410000
QY U00096_34 3400001 3510000
DB U00096_35 3500001 3610000
QY U00096_36 3600001 3710000
DB U00096_37 3700001 3810000
QY U00096_38 3800001 3910000
DB U00096_39 3900001 4010000
QY U00096_40 4000001 4110000
DB U00096_41 4100001 4210000
QY U00096_42 4200001 4310000
DB U00096_43 4300001 4410000
QY U00096_44 4400001 4510000
DB U00096_45 4500001 4610000
QY U00096_46 4600001 463675
DB U00096_46 4600001 463675

```

Continuation (47 of 47) of U00096 from base 4600001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 927; DB 1; Length 39675;
Best Local Similarity 100.0%; Pred. No. 2,3e-264;
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTCATGTTACGCCGATCATGTTAATTTGACGATGATCAGGACAGTCAAGGACTTTGT 60
DB 38429 GTCATGTTACGCCGATCATGTTAATTTGACGATGATCAGGACAGTCAAGGACTTTGT 38370
QY 61 ACTTCGTTTGGATTACTTGGCAATTTAGTTAGCAAAATGACAGCCCGCAGATTCT 120
DB 38369 ACTTCGTTTGGATTACTTGGCAATTTAGTTAGCAAAATGACAGCCCGCAGATTCT 38310
QY 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
DB 38309 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 38250
QY 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAAAATGCATGAGTCTCTCTGAATATGA 240
DB 38249 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAAAATGCATGAGTCTCTCTGAATATGA 38190
QY 241 CATCAACTGTGATATGATATCAATCTGCGGGTAAAGAGAGGTTCTTGTAGCGCG 300
DB 38189 CATCAACTGTGATATGATATCAATCTGCGGGTAAAGAGAGGTTCTTGTAGCGCG 38130
QY 301 TGAATCGCGGAGAGCGGAATGTTGCGTTGATGTTCTGACTGCGCGTGAACAAGAGT 360
DB 38129 TGAATCGCGGAGAGCGGAATGTTGCGTTGATGTTCTGACTGCGCGTGAACAAGAGT 38070
QY 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATACATCAACAAACGGTCAACC 420

```

```

Db      38069 CCAATAAATTCCTGGCCCTGAATTCGGTGCAGATGACTACATACCAACCCGTTCAACC 38010
Qy      421 GCGTGAACGACGATTCGTGCACGCAACCTACTGTCCCGTACCATGATCGGTACTGT 480
Db      38009 GCGTGAACGACGATTCGTGCACGCAACCTACTGTCCCGTACCATGATCGGTACTGT 37950
Qy      481 CAGCGAAGAACGTCGATAGCGTTGAAAGCTCAAGTTCAATGGTTGGAACTGGACATGCA 540
Db      37949 CAGCGAAGAACGTCGATAGCGTTGAAAGCTCAAGTTCAAGTTGGGAACTGGACATGCA 37890
Qy      541 CAGCGCTGTGATGATGGCCCTGATGGCGAGCATCAAGCTGCCGCGAGCGAGATTCCG 600
Db      37889 CAGCGCTGTGATGATGGCCCTGATGGCGAGCATCAAGCTGCCGCGAGCGAGATTCCG 37830
Qy      601 CCGCATGCTTCACTTCTGTGAAAACCGAGCAAAATTCAGTCCCGTGTCACTGTGAA 660
Db      37829 CCGCATGCTTCACTTCTGTGAAAACCGAGCAAAATTCAGTCCCGTGTGAACTGTGAA 37770
Qy      661 GAAATGACCGCGCGGTGAGCTGAAACCGCAGACCGTACTGTGACGTGACGATCCGCCG 720
Db      37769 GAAATGACCGCGCGGTGAGCTGAAACCGCAGACCGTACTGTGACGTGACGATCCGCCG 37710
Qy      721 TATTGTAACATTTGGAATCTACGCGGATAGCGCGGAAATGATGGCCACCATTCACG 780
Db      37709 TATTGTAACATTTGGAATCTACGCGGATAGCGCGGAAATGATGGCCACCATTCACG 37650
Qy      781 TGAAGGTATCGCTTCTGCGGTGATCGAAGATTAATCGGCTTACACCGTCAAAAA 840
Db      37649 TGAAGGTATCGCTTCTGCGGTGATCGAAGATTAATCGGCTTACACCGTCAAAAA 37590
Qy      841 AACGGCGCTTTTAAAGCGCGCTTTTATTTTCAACCTTATTCAGATACGATCTCATC 900
Db      37589 AACGGCGCTTTTAAAGCGCGCTTTTATTTTCAACCTTATTCAGATACGATCTCATC 37530
Qy      901 GTCCGTTGACTCTTACTGCGCTT 927
Db      37529 GTCCGTTGACTCTTACTGCGCTT 37503

```

```

RESULT 3
ECOW93/c 338534 bp DNA linear BCT 30-JAN-2001
DEFINITION Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.
ACCESSION U14003
VERSION U14003.1 GI:1263172
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 338534)
AUTHORS Blattner, F.R.
          Burland, V., Plunkett, G. III, Sofia, H.J., Daniels, D.L. and
          Blattner, F.R.
TITLE Analysis of the Escherichia coli genome VI: DNA sequence of the
JOURNAL region from 92.8 through 100 minutes
MEDLINE Nucleic Acids Res. 23 (12), 2105-2119 (1995)
PUBMED 7610040
REFERENCE 2 (bases 1 to 338534)
AUTHORS Plunkett, G. I. I.
TITLE Direct Subcloning
JOURNAL Submitted (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
COMMENT On Apr 17, 1996 this sequence version replaced gi:536929.
          This sequence was determined as part of the E. coli Genome Project
          (Frederick R. Blattner, director) at the University of
          Wisconsin-Madison. Supported by award HG00301 from the NIH Human
          Genome Project. The entire sequence was independently determined
          from E. coli MG1655; overlaps with other sequence determinations
          are annotated. The start of this entry overlaps the end of the
          entry ECOW89 (U00006) by 1885 bp.

```

```

FEATURES
Source
Location/Qualifiers
1..338534
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/map="92.8 to 100 minutes"
/note="This sequence comprises the following lambda
clones: DD928(EC27-278), DD930(EC30-188),
DD933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070),
DD941(EC17-142), DD945(EC17-8), DD947(EC248-34),
DD949(EC22-169), DD952(EC27-297), DD953(EC27-409),
DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2),
DD962(EC21-104), DD965(EC23A-40), DD968(EC30K60A-49p),
DD970(EC19-202), DD974(EC19-61), DD975(EC18-233),
DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3),
DD984(EC17-101), DD987(EC18-115), DD990(EC17-136),
DD992(EC18-282), DD995(EC17-115), DD997(EC18-113); M13mp19
or Janus vectors were used for subcloning"
1..1885
/note="1885 bp overlap with end of Genbank Accession
Number U00006 (ECOW89)"
complement(1..398)
/note="ORF_f326 of Genbank Accession Number U00006
(ECOW89)"
/codon_start=1
/label_table=11
/protein_id="AA96985.1"
/db_xref="GI:1263173"
/translation="MGFTTRVKSSEASEKKRPFNFALFMDKRYGFFILAIIVAFGSLSP
EYFLTNNTQIFVQSSVTVLIGMGEFFALIVAGIDLVSQALIALSGMTAKKLMLAGV
DPEFLAMIGVVLVGALNGCLVWWTGL"
277..397
/note="predicted bend of 75.10 degrees"
complement(377..1909)
/note="ORF_f510"
/codon_start=1
/label_table=11
/protein_id="AA96986.1"
/db_xref="GI:536931"
/translation="MATPYISMAGIGCKSPFPAHALKSVNLTVYPGEIHALLGENGAK
STLMKVLGSIHEPTKGTITINNIYVNDKHLKLAAGLGIIYQELSVIDELVLEMLY
IGRHLYTKIIGVNIIDREMRVRAAMLRVGLKVDLDEKVNLSISHROMELIATL
MDAYVIMIDREPTSLTNKEVDYLPLNOLKEGATYVISHKLAETIRIDCRVYVM
KDGSSVCSGIVSDVSHDDIVRLMVGRELONRPMKENVSNLAHETVFEVRVNTSRDR
KVVRDLSFVSCGELIGFAGLVSGRTLEMLNCLFVGRKAGEIRLNGKDLSFRSLD
AVKQGAAYTTESRNDGFPNFSIAQNVAAISSLKDQGXKAGMLFHEVDERTAEQD
RELALKCHSVNQNTIELSGNQKVLISKMLCCPEVLIIPDEPTGIDVGAKAETIK
VMROLADQGVILMVSSELPEITTCDRIVAVCEGLITQILNTRDMSESEIIMVAMP
QE"
809..931
/note="predicted bend of 74.75 degrees"
1768..1889
/note="predicted bend of 75.87 degrees"
complement(1983..2010)
/note="promoter matrix score of 55; putative; resembles
flh-regulated promoter"
complement(1985..2021)
/standard_name="RBP; repetitive extragenic palindromic
element"
/note="contains 1 REP sequence"
complement(2036..2971)
/note="ORF_f311"
/codon_start=1
/label_table=11
/protein_id="AA96987.1"
/db_xref="GI:536932"
/translation="MKKYLKRYSGTLVGLMSTSAFPAALAVVLTLSNPFPMVKKK
GIEDEAKTVLSVDIFASPSGDPOSQOLFEDLSNKNKYGIAFAPLSVSNLVMPVAR

```

```
AMKGIYLVNIDKIDMDNLKKAGNVEAFVTTDNVAVGAKGASPIIDLGAGEGEVA
IIEGKAGNAGEARNGATEAPFKKASQIKLIVASOPADMDRIKALDVATNTVLRNPNIK
AIYCAANDTAMGAOVAANAGKTGKVLVCTGCTGIPARKNVAGGQMTATVQNPADIG
ATGKLMDVDEKSGKVIPLDKABEFVLVSIIIVTQ"
misc_structure
3010..3121
/!note="predicted bend of 73.22 degrees"
CDS
complement (3030..3953)
/!gene="ORF_f307"
/codon_start=1
/transl_table=11
/!label=ORF_f307
/!protein_id="AAA96988.1"
/!db_xref="GI:536933"
/!translation="MASSIQGQENVMSQSEPDALPNGIGLAPYLAKMKOBGMTENESR
IYEWLKPENLSCAPAIKDVAVALAVSEAMIVKSLGIFSGFNRILRSLADPVSQSE
OVLPSBELADEVAPQDVNVKENVITLRTIMEGOSIVNDEIHRARAFQOARQRLTGA
GGSNICADVQHKFLRIGVRCOAYPDAMMSASLLOEGDVIVTVTHSGTSVKA
VELAKNGAKTICITHSYHSPITKADYIICSRAETPLGRNMSARILQVTLIDAF
VSAQNTIEQANINMOKTGATVDPFSPQALK"
complement (3952..3979)
promoter
/!note="promoter matrix score of 61; putative"
misc_structure
4151..4251
/!note="predicted bend of 72.06 degrees"
gene
4279..4728
/!gene="yjcA"
/!codon_start=1
/transl_table=11
/!label=ORF_O149
/!protein_id="AAA96989.1"
/!db_xref="GI:536934"
/!translation="MKKIAFCCHVIFILKHEIVAHMERGVVIDKGTWSERTDYP
HYASQVALAVAGGVDGIIICCTGCGEISTAAKNGFIRAVVCSSEYSQOLSRQNDT
NYLAFGRVIGLEIAKMTVDMAQYEGGRHOQREVAITAIQRORN"
complement (4520..16191)
misc_feature
/!note="corresponds to Genbank Accession Number D90227
(ECOPHN) 1..11672; the related sequence Genbank Accession
Number J05260 (ECOPHNAO) is from E. coli B, not K-12"
complement (4739..5146)
gene
/!gene="phnQ"
complement (4739..5146)
CDS
/!gene="phnQ"
/!note="annotated in Genbank Accession Number J05260
(ECOPHNAO), but probably not a gene"
/codon_start=1
/transl_table=11
/!protein_id="AAA96990.1"
/!db_xref="GI:536935"
/!translation="MPAPYTSPPILIFHSPLFIQLPTLLIKALAIYOLPAATFKL
AAAVLILAPIAVFTLPAVARIFVSPFTVAVIVIRNLINANGVVDNIGLRNVD
SVSRVKNRSGKRFISQPKETLSTLIITHPKK"
complement (5272..6030)
gene
/!gene="phnP"
complement (5272..6030)
CDS
/!gene="phnP"
complement (5272..6030)
/!codon_start=1
/transl_table=11
/!protein_id="AAA96991.1"
/!db_xref="GI:536936"
/!translation="MSLTLTLTGAGAGVPAKNGCECAACARARSPOYRRQSGGVV
KENDAITLIDAGLHDLDADRWSPGFOQLTLTHMDHVOGLFPLRWGQDPIPYGSP
DEQGGDILFKHGLDPSHTYBPFVVFPLQGLQVPLPLNLSKLTFGLLIETASRYA
WLSPTNAGIPEKTLKFLRNNOQVAVVMDCSHPRADAPRNHCDLNTVIALNQVIRSPRV
ILTHISHOPDLMENALPSGFVEVFDGMEIGVA"
5452..5586
misc_structure
/!note="predicted bend of 75.60 degrees"
gene
complement (6032..6466)
/!gene="phnO"
complement (6032..6466)
CDS
/!codon_start=1
/transl_table=11
```

```
/!protein_id="AAA96992.1"
/!db_xref="GI:536937"
/!translation="MPACELRPATQYTDVAVALICELKQAEFDHRAFRVGNANLRD
PNMRYHLADLSDGVVMIGLHLQFHLHVNVMGELOELVVMQOARGLVNGSKLMAAE
BEARQGAENTELSTVNRKRDHARFYLREBYSRHSRFTKAL"
complement (6453..7010)
gene
/!gene="phnM"
complement (6453..7010)
CDS
/!gene="phnM"
complement (6453..7010)
/!note="HisP-like nucleotide binding protein (ATP-binding
protein)"
/!codon_start=1
/transl_table=11
/!protein_id="AAA96993.1"
/!db_xref="GI:536938"
/!translation="MMGKLIIMGSPSGSKSLAEILRROTOLLVAHRYITDASA
GSENHIALSEOFPTFRAGONLILASRHNAGLITGVGEIIDLHAGFDVLVNGSRHL
PQARARQOSALIPVCLQVSPETLRORENRERENASBINARLARARYTPDCHTLNN
DGLSQSVDTLTLIHQEKHKL"
complement (7010..8146)
gene
/!gene="phnM"
complement (7010..8146)
CDS
Query Match 100.0%; Score 927; DB 1; Length 338534;
Best Local Similarity 100.0%; Pred. No. 3.5e-264;
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATGTTACGCCGATCATGTTAATTGTCAGACATGATCAGGAGGTGAGGACTTTTGT 60
DB 331239 GTCATGTTACGCCGATCATGTTAATTGTCAGACATGATCAGGAGGTGAGGACTTTTGT 331180
QY 61 ACTTCGTTTCGATTTAGTTAGTTGCGAATTTAGTACGAAACATGACAGCCCGACATTC 120
DB 331179 ACTTCGTTTCGATTTAGTTAGTTGCGAATTTAGTACGAAACATGACAGCCCGACATTC 331120
QY 121 TATCGTTGAGACGAGTTGGTGAACACGACCAACAGTTGAAAAGTATTTGGAACGGAAAG 180
DB 331119 TATCGTTGAGACGAGTTGGTGAACACGACCAACAGTTGAAAAGTATTTGGAACGGAAAG 331060
QY 181 CTATGATGTTTTGCAACGACGACATGCGCGGAAATGATCAGATCCTCTGGAATATGA 240
DB 331059 CTATGATGTTTTGCAACGACGACATGCGCGGAAATGATCAGATCCTCTGGAATATGA 331000
QY 241 CATCAACCTGATGATCATGATATCATGCGCGGTAAGAAAGTCTTCTGTACCGCG 300
DB 330999 CATCAACCTGATGATCATGATATCATGCGCGGTAAGAAAGTCTTCTGTACCGCG 330940
QY 301 TGAACCTGCGGACGACGAGATGTTGCTGATGTTCTGACTGCGCGTGAACAGAACT 360
DB 330939 TGAACCTGCGGACGACGAGATGTTGCTGATGTTCTGACTGCGCGTGAACAGAACT 330880
QY 361 CGATTAATAATCTCGGCGCTGGAATCGGTGACATGATCATCAACCAACCGTTCAACCC 420
DB 330879 CGATTAATAATCTCGGCGCTGGAATCGGTGACATGATCATCAACCAACCGTTCAACCC 330820
QY 421 GCGTGAATCGACGATTCGTGAACGCAACCTACTGTCCTGACATGAAATCTGGTACTGT 480
DB 330819 GCGTGAATCGACGATTCGTGAACGCAACCTACTGTCCTGACATGAAATCTGGTACTGT 330760
QY 481 CAGCGAAGAACGTTCGAGCGTTGAAAGCTACAAAGTTCAATGTTGGAACTGACATCA 540
DB 330759 CAGCGAAGAACGTTCGAGCGTTGAAAGCTACAAAGTTCAATGTTGGAACTGACATCA 330700
QY 541 CAGCGTTCGTTGATCGGCGCTGATGAGGACGACAGTCAAGCGCGCGGACGAGTCCG 600
DB 330699 CAGCGTTCGTTGATCGGCGCTGATGAGGACGACAGTCAAGCGCGCGGACGAGTCCG 330640
QY 601 CGCGATGCTTCACTCTGTGAAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGAAA 660
DB 330639 CGCGATGCTTCACTCTGTGAAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGAAA 330580
QY 661 GAAAATGACCGGCGTGAATGAAACCGACGACCGTACTGTGACGATCGCGCG 720
```

```
Db      330579 GAATAAGACCGCGCGTGAAGTGAACCGACGACCGTACTGTAGACGTGACATCCGCCG 330520
Qy      721 TATTCCTAAACATTTGCAATCTACGCGCGATAGCCGGAATATCATCGCCATTCACGG 780
Db      330519 TATTCCTAAACATTTGCAATCTACGCGCGATAGCCGGAATATCATCGCCATTCACGG 330460
Qy      781 TGAAGGTATTCGCTTTCGCGGTATGTGAAAGTAAATCGGCTTTACCAACGTCGCAAAAA 840
Db      330459 TGAAGGTATTCGCTTTCGCGGTATGTGAAAGTAAATCGGCTTTACCAACGTCGCAAAAA 330400
Qy      841 AACGGGCTTTTAGGCGCCGTTTATTTTCAACCTATTTCCAGATAGTAATCATC 900
Db      330399 AACGGGCTTTTAGGCGCGCTTTTATTTTCAACCTATTTCCAGATAGTAATCATC 330340
Qy      901 GTCGGTGTACTTCTTACTGCGCTT 927
Db      330339 GTCGGTGTACTTCTTACTGCGCTT 330313

RESULT 4
LOCUS   AE005671/c
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 290
ACCESSION AE005671
VERSION   AE005671.1 GI:12519428
KEYWORDS
SOURCE
ORGANISM  Escherichia coli O157:H7 EDL933
REFERENCE 1 (bases 1 to 6794)
AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
TITLE      Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL   Nature 409 (6819), 529-533 (2001)
MEDLINE   21074935
PUBMED    11206551
AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
TITLE      Direct Submission
JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
     source
         1..6794
         /organism="Escherichia coli O157:H7 EDL933"
         /mol_type="genomic DNA"
         /strain="EDL933"
         /serotype="O157:H7"
         /db_xref="taxon:155864"
         /note="enterohaemorrhagic"
     gene
         136..609
         /gene="creA"
         /note="synonym: Z6000"
         136..609
         /gene="creA"
         /note="synonym: Z6000"
     CDS
         136..609
         /function="orf; Unknown function"
         /note="Residues 1 to 157 of 157 are 100.00 pct identical
         to residues 1 to 157 of 157 from Escherichia coli K-12
         strain MG1655: B4397"
         /codon_start=1
         /transl_table=11
         /product="orf, hypothetical protein"
         /protein_id="AAG59577.1"

     gene
         /db_xref="GI:12519429"
         /translation="MKYKHLILSLIMLGPLAAHEIGSVDTVFEMKIGPDHKLIVEA
         FDDPDVKNVTCVSVRAKGTGIGLGLADTSPAAISCCQVGPILSDRIKNAQGE
         VWEKRTSLVPSGIQVRRFYDAKRNALAVIASDKVESPKNIAISAVPMWRQ"
         622..1311
         /gene="creB"
         /note="synonym: Z6001"
         622..1311
         /gene="creB"
         /function="regulator; Global regulatory functions"
         /note="Residues 1 to 229 of 229 are 98.25 pct identical to
         residues 1 to 229 of 229 from Escherichia coli K-12 Strain
         MG1655: B4398"
         /codon_start=1
         /transl_table=11
         /product="catabolic regulation response regulator"
         /protein_id="AAG59578.1"
         /db_xref="GI:12519430"
         /translation="MQRTVWLVEDEQIADTLVYMLQEGFAVEVBERGLPYLDKAR
         QQVDPVMIADVGLPDISGFEELCOLALHPALPVLTARSEEDVRLGLEICADYV
         AKFSPREVCARVRLTRKPKSTSPVYRIGHFELNEPPAQISWDFPLTRYEF
         LLKTLTKSGPRVWSRQQLMDIWMEDAQDTYRTVTHTI KTLRAKLRAPNDPSPINT
         HRGMGYSIRSL"
         1311..2735
         /gene="creC"
         /note="synonym: Z6002"
         1311..2735
         /gene="creC"
         /function="enzyme; Global regulatory functions"
         /note="Residues 1 to 474 of 474 are 98.52 pct identical to
         residues 1 to 474 of 474 from Escherichia coli K-12 Strain
         MG1655: B4399"
         /codon_start=1
         /transl_table=11
         /product="catabolite repression sensor kinase for Phob;
         alternative sensor for pho regulon"
         /protein_id="AAG59579.1"
         /db_xref="GI:12519431"
         /translation="NRIGKRLILGYFLVAVAAFPVIAIFVKEKRGVRRATEGTLID
         TLTILAELARPDLGSDPHTGOLAQAFNLOLRPPANIGIKVNEVYVNTDAG
         KTLFDSANKAVGDDYSRNDWMTTLRGQYGASTLPNDPADPRESSVWYVAPINDDSRD
         IGVLSVGKPNAAAPVIXSERITLWASAILGLIALVIGKXWVWVNRISARLTRAD
         SVTDNKPVPVPEIGSSBELKLAQLESMSKVEKGVYIQVYVALTHELKSPLVALRG
         AAELIEEGPPEPVAFRTDNLITLONARMOALVETLRLORLENRQEVVLTAVALRG
         RRVSEARVTLAKYLDNCGI PDVALSRIPFRFVSLPRANQKSSGLGLAFVSEARLEN
         AEVDQHVNLAKYLDNCGI PDVALSRIPFRFVSLPRANQKSSGLGLAFVSEARLEN
         GEVTLNRVQEGVGLASLRHRIPT"
         2793..4145
         /gene="creD"
         /note="synonym: Z6003"
         2793..4145
         /gene="creD"
         /function="putative membrane; Colicin-related functions"
         /note="Residues 1 to 450 of 450 are 98.22 pct identical to
         residues 1 to 450 of 450 from Escherichia coli K-12 Strain
         MG1655: B4400"
         /codon_start=1
         /transl_table=11
         /product="tolerance to colicin E2"
         /protein_id="AAG59580.1"
         /db_xref="GI:12519432"
         /translation="MLKSPLEFWKMTTLFGAVLLILLIPILIRQVIVERADYRSVEDA
         IROSTSGPQKLVGPLAIPTVELTYTQEDKATYERKSFHFHFLPESLVADGQNVBE
         RRTGIYTGQVMSDGLTLKADPVSRLSDAPRTGKXPITVSVGDAGIGVAKPAE
         VNGTALTTPRGICIEGGCGGVNHPLEBGRMKNTKLNALNLSTGDDISVYVAGNS
         ENTLTISNWPSPFLGDFLPKAKREVSSGFQAQOQSWFANNIGERFASGNDTGWENFP
         AFSVAATTPADQVQLTRATKAIILALITLFAAFVFEETLTQRHLHPKQYLLVGLSLV
         MEYLIALTASEHGFTVAMIIASLIGALNNGYILQVFLGKMRNSMLFTLALLLDGVV
         WGLINADSAALLLGTSVLVVALAGMVFVRNIDWVAFSLPKMKASKEVTTDDELRLWK
         "
         complement(4205..4921)
         /gene="arca"
         /note="synonym: Z6004"
```

CDS
 complement (4205..4921)
 /gene="arcs"
 /function="regulator; Global regulatory functions"
 /note="Residues 1 to 238 of 238 are 100.00 pct identical to residues 1 to 238 of 238 from Escherichia coli K-12 Strain MG1655: B4401"
 /codon_start=1
 /transl_table=11
 /product="negative response regulator of genes in aerobic pathways" (sensor, ArcB and CpxA)"
 /protein_id="AAG59581.1"
 /db_xref="GI:12519433"
 /translation="MOTPHILIVEDELVTRNTLKSIFPAEGYVFEATDGAEMHOIIS
 EYDINLVMDINLPGKNGLLAREDAQVALMFLTGDRNEVDKIGLEIADYIT
 KPNPRELITLRARNLISRTMNLGTVSEERSVSEYKFKMGWELDINSRELIGPDEQIK
 LRSSEFRMLHFCENPGKIQSRALKKMTGRELKPHRITVDVITIRIRKAFESTPDT
 PEIATIHGEGRFCGDLSD"
 5017..5157
 /gene="yjjy"
 /note="synonym: Z6005"
 5017..5157
 /gene="yjjy"
 /function="orf; Unknown function"
 /note="Residues 1 to 46 of 46 are 100.00 pct identical to residues 1 to 46 of 46 from Escherichia coli K-12 Strain MG1655: B4402"
 /codon_start=1
 /transl_table=11
 /product="orf; hypothetical protein"
 /protein_id="AAG59583.1"
 /db_xref="GI:12519434"
 /translation="MTKVRNCVLDALSIINNNIISLVGTFPODPVSKTAVILITLT
 AT"
 5557..6243
 /gene="laet"
 /note="synonym: Z6006"
 5557..6243
 /gene="laet"
 /function="orf; Unknown function"
 /note="Residues 1 to 228 of 228 are 97.80 pct identical to residues 1 to 228 of 228 from Escherichia coli K-12 Strain MG1655: B4403"
 /codon_start=1
 /transl_table=11
 /product="orf; hypothetical protein"
 /protein_id="AAG59583.1"
 /db_xref="GI:12519435"
 /translation="MRITIIIVAPARAENITGAADAMKTGSELRIVDSQAHLEPAT
 RMVHSGSDIINDIKVFTPLASLHDVDTVAFTARSRAKHVYATPELVLEKES
 SMMSHAALVGRBDSGLTNEELADLVDPVPAADYSLNKGAVXYCYOLAILIG
 OPTKSDTADQHQLOALREBRAMALLITLAVADDIKLVDWLQQRIGLLEQRDTAMHLRL
 LHDIEKNITIK"
 ORIGIN
 Query Match 99.3%; Score 920.6; DB 1; Length 6794;
 Best Local Similarity 99.6%; Pred. No. 1.3e-262;
 Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GTCAATTAACGCCGATCATATTAATTTGACATGATCAGGACAGGTAGGACATTTGT 60
 DB 5021 GTCAATTAACGCCGATCATATTAATTTGACATGATCAGGACAGGTAGGACATTTGT 4962
 QY 61 ACTTCGTGTTGATTTAGTTGACATTTAGGTAGCAAAATGACACCCCGACATTTCT 120
 DB 4961 ACTTCGTGTTGATTTAGTTGACATTTAGGTAGCAAAATGACACCCCGACATTTCT 4902
 QY 121 TATCGTTGAAGAGAGTTGTATACACGACACGTTGAAAGTATTTTCCAAAGCGAAG 180
 DB 4901 TATCGTTGAAGAGAGTTGTATACACGACACGTTGAAAGTATTTTCCAAAGCGAAG 4842
 QY 181 CTATGATTTTTCGAAGGACAGATGCGCGGAAATCATCATGATCTCTCTGAATATGA 240
 DB 4841 CTATGATTTTTCGAAGGACAGATGCGCGGAAATCATCATGATCTCTCTGAATATGA 4782

QY 241 CATCAACCTGTGATCATATGATATCAATCTGCCGGTAAAGACGATCTTCTGTTAGCGG 300
 DB 4781 CATCAACCTGTGATCATATGATATCAATCTGCCGGTAAAGACGATCTTCTGTTAGCGG 4722
 QY 301 TGAATCTGCGGACGACGCAATGTTGATGTTTCCGATCGGCGCGGACAAAGT 360
 DB 4721 TGAATCTGCGGACGACGCAATGTTGATGTTTCCGATCGGCGCGGACAAAGT 4662
 QY 361 CGATTAATTTCTGCGGCTGAAATCGTGCAATGATCATACATCAACCGTTAACCC 420
 DB 4661 CGATTAATTTCTGCGGCTGAAATCGTGCAATGATCATACATCAACCGTTAACCC 4602
 QY 421 GCGTGAATGACGATCTGTCACGCAACCTAATGTCGCGTACCAAGTATTTGGTACTGT 480
 DB 4601 GCGTGAATGACGATCTGTCACGCAACCTAATGTCGCGTACCAAGTATTTGGTACTGT 4542
 QY 481 CAGCGAAGACGTCGTAAGCTTGAAGCTACAAATGATGTTGGAATGACATGAA 540
 DB 4541 CAGCGAAGACGTCGTAAGCTTGAAGCTACAAATGATGTTGGAATGACATGAA 4482
 QY 541 CAGCGTGTGATCGGCGGCGGATGAGGACGATGACGATGACGCGGCGGATGTCG 600
 DB 4481 CAGCGTGTGATCGGCGGCGGATGAGGACGATGACGATGACGCGGCGGATGTCG 4422
 QY 601 CGCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGATGATGCTGAA 660
 DB 4421 CGCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGATGATGCTGAA 4362
 QY 661 GAAATGACCGGCGGCGGATGAGGACGCAACCGATCTGTAAGCTGATCGGCGG 720
 DB 4361 GAAATGACCGGCGGCGGATGAGGACGCAACCGATCTGTAAGCTGATCGGCGG 4302
 QY 721 TATTCGTAACATTTGCAATCTACGCGGATGACGCGGAAATCATGCGCAATTCACG 780
 DB 4301 TATTCGTAACATTTGCAATCTACGCGGATGACGCGGAAATCATGCGCAATTCACG 4242
 QY 781 TGAAGTTATGCTTCTGCGGCGGATGATGGAAGATTAATGCGCTTTCACCGTCAAAAA 840
 DB 4241 TGAAGTTATGCTTCTGCGGCGGATGATGGAAGATTAATGCGCTTTCACCGTCAAAAA 4182
 QY 841 AACGGGCTTTTGAACGCGCTTTTATTTTCAACCTTATTTCCAGATCGTAACTGATC 900
 DB 4181 AACGGGCTTTTGAACGCGCTTTTATTTTCAACCTTATTTCCAGATCGTAACTGATC 4122
 QY 901 GTCCGTGTAACTTCTTAACTGCTTT 927
 DB 4121 GTCCGTGTAACTTCTTAACTGCTTT 4095

RESULT 5
 AP002569/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCES
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE

AP002569 168764 bp DNA linear BCT 20-MAR-2004
 Escherichia coli O157:H7 DNA, complete genome, section 20/20.
 AP002569 BAO00007
 AP002569.1 GI:13364704
 Escherichia coli O157:H7
 Escherichia coli O157:H7
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1 Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
 Kurokawa, K., Ishii, K., Hattori, M., Tatsu, I., Abe, H., Iida, T.,
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
 Sasaki, C. and Shinagawa, H.
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 gene of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak
 Genes Genet. Syst. 74 (5), 227-239 (1999)
 20198780
 10734605

AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurakawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shingawa,H. and
Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE 20557356
PUBMED 11108008
REFERENCE
AUTHORS 3
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutemoto,C.H., Kurakawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,C.K., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,K., and
Shingawa,H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
JOURNAL Gene 258 (1-2), 127-139 (2000)
MEDLINE 20564182
PUBMED 1111050
REFERENCE
AUTHORS 4
Hayashi,T., Makino,K., Ohnishi,M., Kurakawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,W., Itoe,T., Iida,T., Takami,H., Honda,T., Sasaki,K.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shingawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156231
PUBMED 11258796
REFERENCE
AUTHORS 5 (bases 1 to 168764)
Hattori,M., Ishii,K. and Shiba,T.
TITLE Direct Subinjection
Submitted (26-JUN-2000) Masahira Hattori, Kitasato Institute for
Life Sciences, Kitasato University; Kitasato 1-15-1, Sagamihara,
Kanagawa 228-8555, Japan
(E-mail:hattori@genome.1s.kitasato-u.ac.jp,
URL: http://genome.1s.kitasato-u.ac.jp/, Tel:81-42-778-8194,
Fax:81-42-778-8193)
COMMENT genome project
This work was done in collaboration with Tetsuya Hayashi, Makoto
Ohnishi, Keisuke Nakayama (Miyazaki Medical College), Kozo Makino,
Ken Kurakawa, Katsushi Yokoyama, Masashi Tanaka, Takeshi Honda,
Tetsu Yasunaga, Hideo Shingawa (Osaka University), Takahiro Murata
(Shinshu University), Chang-Gyun Han, Eiichi Ohtsubo, Toru Tobe,
Chihiro Sasaki (University of Tokyo), Hideko Takami (Japan Marine
Science and Technology Center), Satoru Kuhara (Kiyushu University), and
supported by the Research for the Future Program of the Japan
Society for the Promotion of Science.
FEATURES
source Location/Qualifiers
1.168764
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..243
/gene="EC65227"
79..243
/gene="EC65227"
/note="identical to B4250_ECOLI g1|1790699 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38650.1"
/db_xref="GI:13364705"
/translation="MLSVILKLLIILVTPRASTRLLAVRLGLVYSHNMSTVKLFR
LSIFPMIVR"
315..569
gene

CDS
/gene="EC65228"
315..569
/gene="EC65228"
/note="similar to YJGJ_ECOLI g1|1790700 percent identity
92 in 84 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38651.1"
/db_xref="GI:13364706"
/translation="WVLESGHSHDPARDIAVYTHAERATIDYIARLHPGACVY
DPMSTVSGLSAARGHSHLEQLCATPALGAEIKITLKE"
705..1166
/gene="EC65229"
705..1166
/gene="EC65229"
/note="similar to YJGK_ECOLI g1|1790701 percent identity
98 in 153 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38652.1"
/db_xref="GI:13364707"
/translation="MEKMIIGNIHNLOPWLPELRQAIEHKAHTAETPKGHDI
NRLFYLSIDMEPEYARAEYARVLDIOIVLKGQEGMTFTSPAGTPTDMLADK
IAFLPEGVDEKTYILNEDGFVYVPEGVHKLPCAVGAPAQVRAVVMMA"
1289..3061
/gene="EC65230"
1289..3061
/gene="EC65230"
/note="similar to part of (13-616 in 616 aa) YJGL_ECOLI
g1|1790702 percent identity 82 in 590 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38653.1"
/db_xref="GI:13364708"
/translation="NSKISDLNYSCHITLADNPKQSEVLTNRVGNMPPARAGOD
NTRNINPTFPLFVLTITLGYVDSKRSNEGKRWMAHIBSSYTKINDGSEIMK
VWNPNGERADSPKVIIEISLSTITTMGRQGTALIFQPDSTSTRYKSKSERD
SISHLITNLYVLACYOREANKETARLNNHQNNLOKNNLOKNNLOKNNLOKNN
IOELNNSOELNNSOELNNSOELNNSOELNNSOELNNSOELNNSOELNNSOELN
RHEHLIMHVSSEISYVLGNGILEIDKYLELNDHGDINDKEENINDIMNIS
HILIBSDAKNLTNPVTHSMNTPIELRYNNDVKILMCRKSKMYPFDDSAKIEHAC
SYNININFRDSSKVAETLFFINDKEPYKNSPELOELWKLVVYVNDPLNREKTY
LQRIENNVESLEFNKVPVSLIVNDIFMDFPKNTIMINWYFRLKSYEDEKTYFDK
LGVPNNKSENEIMKQPKDYEEKLNNEELKLRFRMOTILKSEYVNSPFIQORLN
TINPENTLRIAEKFGMKKTITTA"
complement (3118..4122)
/gene="EC65231"
complement (3118..4122)
/gene="EC65231"
/note="similar to ARGF_ECOLI g1|1786469 percent identity
87 in 334 aa, similar to ARGJ_ECOLI g1|1790703 percent
identity 99 in 334 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ornithine carbamoyltransferase 1"
/protein_id="BAB38654.1"
/db_xref="GI:13364709"
/translation="MSGFYHGFILKLDTPPAELNSLLQALAKLRDKSGKEEATLT
GNRIALIFEKDSTTRKCSFEVAAYDGAATVVGPSGQIGHKESIQTARVAGRYVD
GIOYRGYGOEIVETLAEGVPMVNGLTNEFPPTQILADLTMOEHLPGKPNEMTLV
YAGDANNNGNSMLEAALITGLDLRLVAPQACPEALVTECRALQOQNGNTLTED
VAKGVGADPIFYTDVWVSGEAKKEKAEIALRLDYOVNSKMWLTGNEVEKHLCLP
AFHDQDTLGCKMAEERFGHGMGEVTVDFESAASIVPQAEHRMHTITKAVNATLSK
"
4284..4700
/gene="EC65232"
gene

Query	Best Local Similarity	99.3%; Score 920.6; DB 1; Length 168764;	Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db	167518	GTATGATGAGCCGATCATGTTAATTTGGACATGATCAGGACAGTCAAGGACTTTTGT	60
Qy	61	ACTTCTGTTTCGATTAGTTAGTTGGCAATTTAGTAGAGCAACATGACAGACCCCGACATTTCT	120
Db	167458	ACTTCTGTTTCGATTAGTTAGTTGGCAATTTAGTAGAGCAACATGACAGACCCCGACATTTCT	167399
Qy	121	TATCGTTGAAGAGCATGTTGGTAAACAGCAACAGCTTGAAGAGGAGGAGG	180
Db	167398	TATCGTTGAAGAGCATGTTGGTAAACAGCAACAGCTTGAAGAGGAGGAGG	167339
Qy	181	CTATGATGTTTTCGAAGCAGACAGATGAGCGGGAAGAGCATGAGATCTCTGTAATATGA	240
Db	167338	CTATGATGTTTTCGAAGCAGACAGATGAGCGGGAAGAGCATGAGATCTCTGTAATATGA	167279
Qy	241	CATCAACCTGGTATCATGATATCAATCTGCGGGTAAAGAACGGCTTCTGTAGACGG	300
Db	167278	CATCAACCTGGTATCATGATATCAATCTGCGGGTAAAGAACGGCTTCTGTAGACGG	167219
Qy	301	TGAACCTGGCGCGAGCGAGCAATGTTGCGTTGATGTTTCTCGATCTGGCGCGTGAACAAGT	360
Db	167218	TGAACCTGGCGCGAGCGAGCAATGTTGCGTTGATGTTTCTCGATCTGGCGCGTGAACAAGT	167159
Qy	361	CGATTAATTTCTGGCTCTGGAATCGGTGACAGATGACTATCAACCAACCGTTACACC	420
Db	167158	CGATTAATTTCTGGCTCTGGAATCGGTGACAGATGACTATCAACCAACCGTTACACC	167099
Qy	421	GCGTGAACCTGAGATTCGTGACGAGCAACCTACTGTCCTCCGTACCATGAATCTGGGTAAGT	480
Db	167098	GCGTGAACCTGAGATTCGTGACGAGCAACCTACTGTCCTCCGTACCATGAATCTGGGTAAGT	167039
Qy	481	CAGCGAAGAACGTCGTAGCGTTGAAGAGCTAAGTTCAATGTTGGAGAACTGGACATCAA	540
Db	167038	CAGCGAAGAACGTCGTAGCGTTGAAGAGCTAAGTTCAATGTTGGAGAACTGGACATCAA	166979

Qy	544	CAGCGCTTCGTATATCGGCGCCCTATATGCGGACGACGTACAGCTCCCGCGGACGATTCGCG	600
Db	16678	CAGCGCTTCGTATATCGGCGCCCTATATGCGGACGACGTACAGCTCCCGCGGACGATTCGCG	166919
Qy	601	CGCCATGCTTCACCTTCGTGTAAAAACCGAGCAAAATTCAGTCCGCGTGTGAAATCGCGTGA	660
Db	166918	CGCCATGCTTCACCTTCGTGTAAAAACCGAGCAAAATTCAGTCCGCGTGTGAAATCGCGTGA	1668599
Qy	661	GAATAAGACCGCGCGGAGAGCTGAAACCCGACACCGTACTGTATGACGTGACGATTCGCGCG	720
Db	166858	GAATAAGACCGCGCGGAGAGCTGAAACCCGACACCGTACTGTATGACGTGACGATTCGCGCG	166799
Qy	721	TATTCGTAAACATTTGAAATCTACCGCGGATACGCGGAAATCATCGCCACCATTCACGG	780
Db	166798	TATTCGTAAACATTTGAAATCTACCGCGGATACGCGGAAATCATCGCCACCATTCACGG	166739
Qy	781	TGAAGTTATTCGCTTCGCGGAGATTCGTGAAGATTATTCGCTTTACACGCTCAAAAA	840
Db	166738	TGAAGTTATTCGCTTCGCGGAGATTCGTGAAGATTATTCGCTTTACACGCTCAAAAA	166679
Qy	841	AACGCGACTTTTAAAGCGCGCTTTATTTTATTTTGAACCTTATTTCCAGATACGTAATCATC	900
Db	166678	AACGCGACTTTTAAAGCGCGCTTTATTTTATTTTGAACCTTATTTCCAGATACGTAATCATC	166619
Qy	901	GGCGGTGTAATCTTCCTTACTGGCTTT 927	
Db	166618	GGCGGTGTAATCTTCCTTACTGGCTTT 166592	
RESULT 6	AE016772/c	8698 bp DNA linear BCT 24-DEC-2002	
LOCUS	Escherichia coli CFT073 section 18 of 18 of the complete genome.		
DEFINITION	AE016772 AE014075		
ACCESSION	AE016772.1 GI:26111641		
VERSION			
KEYWORDS			
SOURCE	Escherichia coli CFT073		
ORGANISM	Escherichia coli CFT073		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (bases 1 to 8698)		
TITLE	Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P., Raeko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L.T., Domeneberg, M.S., and Blattner, F.R. Extensive Mosiac Structure Revealed by the Complete Genome Sequence of Uropathogenic Escherichia coli		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)		
PUBMED	12471157		
REFERENCE	2 (bases 1 to 8698)		
AUTHORS	Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P., Raeko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L.T., Domeneberg, M.S., and Blattner, F.R. Direct Submission		
TITLE	Submitted (20-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..8698		
SOURCE	/organism="Escherichia coli CFT073"		
	/mol_type="genomic DNA"		
	/strain="CFT073"		
	/db_xref="taxon:199310"		
	complement(152..1495)		
	/gene="gntP"		
	/locus_tag="c5401"		
	complement(152..1495)		
	/gene="gntP"		
	/locus_tag="c5401"		
	/function="transport; Transport of small molecules: carbohydrates, organic acids, alcohols"		
	/note="Glucanase permease 3; Escherichia coli K-12"		
CDS			


```

ortholog: b4321; Escherichia coli O157:H7 ortholog: z5919"
/codon_start=1
/translate_table=11
/product="High-affinity gluconate transporter"
/protein_id="AAN83823.1"
/db_xref="GI:26111642"
/translation="MHVNLINLVYFGIGLMLVNLKFKINSVVALLVAAISVGLMAG
DMLSLHTKAGFGNTLGLBLAIIVFGAVYGLMVDGSAHQIAHTLRLARGLRYQI
SVIIIGLIFGLAMFYEVAFIMLAPLIVIAAEAKIPPLKLAIPVAAATYASLPPO
PGPVAAVNAVYAGDMGVYIYGVLTIPSVYAGLIPKFLGNLERPTSPSLAKDOVD
MNNLPSEGVSLVPLIPALIMSTTIANWLVDPFAVEVNFISGSPILAMFIAWVA
FVLFGTARGHDMQWKNWNAFSAVKSIAWYILIIIGAGVYKQTIIDTIGDTIGLMSH
GNISFYIMKMLITVLIRLATVQGVASAMTPAGIISAAILDPATGOLGVVNPALVAT
AAGSNTLTHINASFWLFRGYFDLSVKTDLTKTGLLELVNSVGLIIVLISWVA"
1835..3019
/gene="uxuA"
/locus_tag="c5402"
1835..3019
/gene="uxuA"
/locus_tag="c5402"
/EC_number="4.2.1.8"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/notes="Escherichia coli K-12 ortholog: b4322; Escherichia
coli O157:H7 ortholog: z5920"
/codon_start=1
/translate_table=11
/product="Mannonate dehydratase"
/protein_id="AAN83824.1"
/db_xref="GI:26111643"
/translation="MEQTRWVGPNDPVSALADYRQAGTGTALHHIPGEVSEE
ILKRAIVEDAGLWMSQTSVPIHEDIKTHGYEQMIANYOQTLENLACQGRITCY
NEMPVLDMTRTDLEVLIPDGSKALRFQJLEFAFEMHILKRPGEADYEELAAAB
RAIYMSDEDEKARLTRNIAGLPGAEFGYTLDOFRKHLKLYKQIDKALENFEVFLA
IIPVAEVCVWMAVHPDDPRRITGCPRIYSTIEDQMWDVTNSANGFTMTGSGY
VRAUNDLVMIKQFGRITYFTHLRSTRENEPTFHEAHLNGVDVMEVKAIVEE
HRRKAEKGEDLIIPRPDGHQMLDLKKKTNPQYSAIGRLKGLAEVGVELAIGRAF
SR"
3100..4560
/gene="uxuB"
/locus_tag="c5403"
3100..4560
/gene="uxuB"
/locus_tag="c5403"
/EC_number="1.1.1.57"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/notes="Escherichia coli K-12 ortholog: b4323; Escherichia
coli O157:H7 ortholog: z5921"
/codon_start=1
/translate_table=11
/product="D-mannonate oxidoreductase"
/protein_id="AAN83825.1"
/db_xref="GI:26111644"
/translation="MTTIYDSNLPVAPSMDSRLRSRIYHLGCGAFRRHQALYTHH
LLESTSDWGCIEVNLPGNDRLVLENLKKOQLYVAEKAESTELKTIIGSKELH
PEIDGEGEGLINMAAPOTAIVSLTVEKGYCADPAAGOLDIANNPLIKHLENTLTA
ATGYIEARLRLREKGLKAFYMSCDNVAENGVAIVYGLAQAARDPOLAAIENNV
TPCTWMDRIVRPATPETTIOETADQGVDPQCIACEPRQVTEENFNGRDWQKV
QAFVADVVPFEMMKLRLMLNGSHSLAYIGLYGETTIADVTNPNRYKAAAFIAWQE
QAPLMSBGTDLNAYATLLIERFSNPSLRHRTWQJAMDSQCLPQRLADPVYLHQN
GGSWRIALGVAGMWRVYQGVDEQGNADIVVDPMMAEPKINQVYGAADRVAKGLLS
GFADLDPQNAFVGAVTAAVQQLCERGARBECAAL"
4775..5548
/gene="uxuR"
/locus_tag="c5404"
4775..5548
/gene="uxuR"
/locus_tag="c5404"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/notes="Escherichia coli K-12 ortholog: b4324; Escherichia
coli O157:H7 ortholog: z5922"

```

```

/codon_start=1
/translate_table=11
/product="Uxu operon trancriptional regulator"
/protein_id="AAN83826.1"
/db_xref="GI:26111645"
/translation="MKSATSAQRPYEVGMIRDLIKTPYNGERLPPEREIAEMID
VTRTVREALIMLEIQLVLRGAGIYVLDSGSSNNASPDANVNDGQFPFLQAKR
QULESNIASFALQAOTREDIYVNRQALQEEERLASAGSSSGDMQFPLAIAEYTH
NSMLVLPFQSQMQNENPMWIOHSHLDSDLYRKEMLGDKHKOILAALIKDARAAKL
ANMOHLENKORLBERSVNDIYFDGYLPDSWPLDKVDA"
5953..6345
/gene="yjiD"
/locus_tag="c5405"
5953..6345
/gene="yjiD"
/locus_tag="c5405"
/notes="Escherichia coli K-12 ortholog: b4326; Escherichia
coli O157:H7 ortholog: z5925"
/codon_start=1
/translate_table=11
/product="Hypothetical protein yjiD"
/protein_id="AAN83827.1"
/db_xref="GI:26111646"
/translation="MRKMRQSVQYTVLPSTGNTNTSLRDSVCRDLFQLFSSPSHP
ILLVSGMPMQGHNSDKLLQSWYCRQLSALLFHPRIALQVNLKEAYCHELAISL
EMMLYHDDPPLTFDLVWQKGSWHRTPQ"
complement(6393..6734)
/locus_tag="c5406"
complement(6393..6734)
/locus_tag="c5406"
/notes="Residues 1 to 113 of 113 are 95.57 pct identical to
residues 424 to 536 of 536 from GenPept.129 :
>b|AAL61895.1| (AF289032) putative
phosphoenolpyruvate-protein phosphoryltransferase ppeI
[Escherichia coli]"
/codon_start=1
/translate_table=11
/product="Hypothetical protein"
/protein_id="AAN83828.1"
/db_xref="GI:26111647"
/translation="MIEVPAVMIADPKLASEVDFPSIGTNDLTQYIMAADRGSTYAK
LVDYCDNAVINAIAWYCAQRNNEIPVSMGEMAGDIQQTARLLTWGIDKLSPSRL
PALKAIRTSH"
complement(6771..7760)
/gene="yjiB"
/locus_tag="c5407"
complement(6771..7760)
/gene="yjiB"
/locus_tag="c5407"
/function="putative regulator"
/notes="Escherichia coli K-12 ortholog: b4327; Escherichia
coli O157:H7 ortholog: z5926"
/codon_start=1
/translate_table=11
/product="Hypothetical trancriptional regulator yjiE"
/protein_id="AAN83829.1"
/db_xref="GI:26111648"
/translation="MKRRKRLSAARDVDCPGGRLTENVYMDDCGAVLIANETKMLYD
FULEKCRNFQSAASRVNSQPAFSRRIRALDQAGVBLFNRQVTPYLQISEQKITHS
QIRHLLOQLEESLIELRGSDYAKRIKIAAHSLSGLPSTISQMPFLPAITBAI
DVEEDVLDKRESCSDCIFSHDEDDLEAFPDHI RLPESQLFPVCADEGHEALFWIAQ
PHEPLNYSRNSYMGRLIRTLTRSELSFSFVPSMSLELLKQVALLDCCGIAFWREY
AIOGRFSGQLVILNRDELVIRIQAVAYMNTMRNPVAFRRFRELRELEIERP"
complement(7747..8919)
/gene="lida"
/locus_tag="c5408"
complement(7747..8919)
/gene="lida"
/locus_tag="c5408"
/EC_number="3.4.19.-"
/function="enzyme; Degradation of proteins, peptides,
glyco"
/notes="Escherichia coli K-12 ortholog: b4328; Escherichia

```

```

coli O157:H7 ortholog: z5927"
/codon_start=1
/translation="MIDYSAAGPTLLIQNALHPEDRIGCDVLINAKKIIAVASNIPS
/product="isoaspartyl dipeptidase"
/protein_id="AA083830.1"
/db_xref="GI:26111649"
/translation="MIDYSAAGPTLLIQNALHPEDRIGCDVLINAKKIIAVASNIPS
DIYDCTVVDLSGQILICPGFIDQVHLIGGGEAGPTTRTPPALSLRLTEAGITSVG
LVGTDISIIRHPSPLAKTLRALNBEGISAMLTGAVHSRRTTGSVEVDVAITDRVIG
VKCAISDRHSAAADVYHLNMAAESRVSGLGKRPVTFPHNGSKRLQVYDILEN
CVPISKLPTHTNRVPLFEQALFARGGITDITSSIDEVPAEAGIARAVQAGIP
LARVTLSDNGSOPPEDEGNLTHIGIAGETTLEIVQVWKDYSFSLDALRPLTS
SVAGFLNLRKKEILLPANDADILVMTPEURLIQVYARGLKLVKQKACVKGFETA"

Query Match      98.4%; Score 912.6; DB 1; Length 86898;
Best Local Similarity 99.0%; Pred. No. 5,3e-260;
Matches 918; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

1  GTCATGTTACGCCGATCATGTTAATTTGACATCATCAGGAGGTGAGGACTTTTGT 60
Db 85652  GTCATGTTACGCCGATCATGTTAATTTGACATCATCAGGAGGTGAGGACTTTTGT 85593
Qy 61  ACTTCTGTTTCGATTTAGTTAGTGGCAATTTAGGTAGCAACATGAGACCCCGACATTCT 120
Db 85592  ACTTCTGTTTCGATTTAGTTAGTGGCAATTTAGGTAGCAACATGAGACCCCGACATTCT 85533
Qy 121  TATCGTTGAAGACGAGTTGTTAACAACGACACCTTGAAAAGTATTTTCGAAGCGGAAG 180
Db 85532  TATCGTTGAAGACGAGTTGTTAACAACGACACCTTGAAAAGTATTTTCGAAGCGGAAG 85473
Qy 181  CTATGATATTTTTCGAAGCAGATGAGCGCGGAATGATCATCATCTCTCTGAATATGA 240
Db 85472  CTATGATATTTTTCGAAGCAGATGAGCGCGGAATGATCATCATCTCTCTGAATATGA 85413
Qy 241  CATCAACCTGGTGAATCATGATATCATCTCCGCGGTGAAGACCGTCTTCTGTTAGCGCG 300
Db 85412  CATCAACCTGGTGAATCATGATATCATCTCCGCGGTGAAGACCGTCTTCTGTTAGCGCG 85353
Qy 301  TGAATCTCGCGAGGACGAGATGTTGGTGTGATGTTCTGACGCGCGGTGACACAGAAAT 360
Db 85352  TGAATCTCGCGAGGACGAGATGTTGGTGTGATGTTCTGACGCGCGGTGACACAGAAAT 85293
Qy 361  CGAATAAATCTCGGCTCGAATTCGCTGACATGATGATCATCATCACCACCGTTCAACC 420
Db 85292  CGAATAAATCTCGGCTCGAATTCGCTGACATGATGATCATCATCACCACCGTTCAACC 85233
Qy 421  GCGTGAATCTGACGATTCGTGACGCAACCTACTGTCCTCGTACATGAAATCTGGTACTGT 480
Db 85232  GCGTGAATCTGACGATTCGTGACGCAACCTACTGTCCTCGTACATGAAATCTGGTACTGT 85173
Qy 481  CAGCGAAGAACGTCGTAAGGCTTGAAGCTTCAAGTTCAATGTTGGAACTGGACATCA 540
Db 85172  CAGCGAAGAACGTCGTAAGGCTTGAAGCTTCAAGTTCAATGTTGGAACTGGACATCA 85113
Qy 541  CAGCCGTTGTTGATGAGCCCTGATGAGGAGACATCAAGTCCGCGCAGCGAGCTTCG 600
Db 85112  CAGCCGTTGTTGATGAGCCCTGATGAGGAGACATCAAGTCCGCGCAGCGAGCTTCG 85053
Qy 601  CCGCATGCTTCACTTCTGTGAAGAACCCAGGCAAAATTCAGTCCGCTGTGAATCTGCTGAA 660
Db 85052  CCGCATGCTTCACTTCTGTGAAGAACCCAGGCAAAATTCAGTCCGCTGTGAATCTGCTGAA 84993
Qy 661  GAAAAATGACCGGCGGAGCTGAAACCCGACGACCTGTAAGCTGTAAGCTGCGCG 720
Db 84992  GAAAAATGACCGGCGGAGCTGAAACCCGACGACCTGTAAGCTGTAAGCTGCGCG 84933
Qy 721  TATTTGTTAAACATTTTGAATCTACCGCCGATACCGCGGAATCATCGCACCATTTACCG 780
Db 84932  TATTTGTTAAACATTTTGAATCTACCGCCGATACCGCGGAATCATCGCACCATTTACCG 84873
Qy 781  TGAAGTTATCGCTTCTCGGCTGATCTGGAAGATTAAATCGGCTTTTACACCGTCAAAAAA 840
Db 84872  TGAAGTTATCGCTTCTCGGCTGATCTGGAAGATTAAATCGGCTTTTACACCGTCAAAAAA 84813

```

```

Qy 841  AACGGCGCTTTTATGAGCGGCTTTTATTTTCACTTAATTTTCAGATACGTAACATC 900
Db 84812  AACGGCGCTTTTATGAGCGGCTTTTATTTTCACTTAATTTTCAGATACGTAACATC 84753
Qy 901  GTCCGTTGTAACCTTTTCTTACTGCGCTT 927
Db 84752  GTCCGTTGTAACCTTTTCTTACTGCGCTT 84726

RESULT 7
ECODYE
LOCUS
DEFINITION
E.coli dye gene coding for Dye protein, complete cds.
VERSION
M10044
M10044.1 GI:145817
KEYWORDS
dye gene.
SOURCE
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 1468)
Drury, L.S. and Buxton, R.S.
DNA sequence analysis of the dye gene of Escherichia coli reveals
amino acid homology between the dye and Ompr proteins
J. Biol. Chem. 260 (7), 4236-4242 (1985)
85157583
2984198
COMMENT
Original source text: E.coli K12 DNA, clone pRB52.
Draft entry and sequence in computer readable form kindly provided
by R.S.Buxton, 01-AUG-1985 [1].
Mutation of the dye gene results in sensitivity to dyes, envelope
protein changes, loss of expression of alkaline phosphatase, and
reduced transcription of sex factor F genes. The positions of the
dye and ompr genes on the K12 map are indicative of gene
duplication and the Dye protein is 28% homologous to the Ompr
protein.
A potential transcription termination site is located at positions
906-940 and a ribosome binding site at 161-166. An UPR was found
on the complementary strand at positions >1542 to 948 (AA at 1541).
FEATURES
source
location/Qualifiers
1..1468
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
98..814
/gene="dye"
98..814
/gene="dye"
1
/codon_start=1
/translation="MIDYSAAGPTLLIQNALHPEDRIGCDVLINAKKIIAVASNIPS
DIYDCTVVDLSGQILICPGFIDQVHLIGGGEAGPTTRTPPALSLRLTEAGITSVG
LVGTDISIIRHPSPLAKTLRALNBEGISAMLTGAVHSRRTTGSVEVDVAITDRVIG
VKCAISDRHSAAADVYHLNMAAESRVSGLGKRPVTFPHNGSKRLQVYDILEN
CVPISKLPTHTNRVPLFEQALFARGGITDITSSIDEVPAEAGIARAVQAGIP
LARVTLSDNGSOPPEDEGNLTHIGIAGETTLEIVQVWKDYSFSLDALRPLTS
SVAGFLNLRKKEILLPANDADILVMTPEURLIQVYARGLKLVKQKACVKGFETA"
ORIGIN
699 bp upstream of AccI site.

Query Match      98.4%; Score 912; DB 1; Length 1468;
Best Local Similarity 99.9%; Pred. No. 3.7e-260;
Matches 923; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

5  TGTACGCCGATCATGTTAATTTGACATCATCAGGAGGTGAGGACTTTTGTACT 63
Db 1  TGTACGCCGATCATGTTAATTTGACATCATCAGGAGGTGAGGACTTTTGTACT 60
Qy 64  TCCTGTTGATTTAGTTAGTGGCAATTTAGTGAAGCAACATGAGACCCCGACATTTCTAT 123
Db 61  TCCTGTTGATTTAGTTAGTGGCAATTTAGTGAAGCAACATGAGACCCCGACATTTCTAT 120
Qy 124  CGTTGAAGACGAGTTGTTAACAACGACGTTGAAAAGTATTTTTCGAAGCGGAAGCTTA 183

```



```
gene
4090..4416
/gene="trpR"
/locus_tag="SF4425"
4090..4416
/gene="trpR"
/locus_tag="SF4425"
/notes="Residues 1 to 108 of 108 are 100 pct identical to
residues 1 to 108 of a 108 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_291006.1"
/codon_start=1
/transl_table=1
/product="regulator for trp operon and arh; trp
aporepressor"
/protein_id="AAN45839.1"
/db_xref="GI:24054962"
/translation="MAQSPYSAPAAAEQHOMLRFVPLDKAYONDIHLPLNMLT
PDEREALGTRVRIVEELIRGEMSORLEKNELDAGATITRGSNSLKAAPVELKOWLE
VLKSD"
complement(4513..5030)
/gene="yjiX"
/locus_tag="SF4426"
/pseudo
complement(4513..5030)
/gene="yjiX"
/locus_tag="SF4426"
/notes="similar to Escherichia coli K12orf, conserved
hypothetical protein gi:1790855 (174 aa). BLAST with
identity of 95% in 172 aa. This CDS contains a frameshift.
The sequence has been checked and is believed to be
correct."
/pseudo
/codon_start=1
/transl_table=1
5073..5720
/gene="gpmB"
/locus_tag="SF4427"
5073..5720
/gene="gpmB"
/locus_tag="SF4427"
/notes="Residues 1 to 215 of 215 are 100 pct identical to
residues 1 to 215 of a 215 aa protein from Escherichia
coli O157:H7 ref: NP_313380.1"
/codon_start=1
/transl_table=1
/product="phosphoglyceromutase 2"
/protein_id="AAN45840.1"
/db_xref="GI:24054963"
/translation="MLQVYLVRHGETONNAERRIGOSDPLTAKGEQAOAQTAK
WRQLVNGTVYDGRIPGESMOELSDRVNALESCHDLPGSRPLVSHGALCVST
ILGLPAAERRLRLRNCISIRVDYQESLWLASGVVETAGDISHADALDELQR"
complement(5717..6586)
/gene="rob"
/locus_tag="SF4428"
complement(5717..6586)
/gene="rob"
/locus_tag="SF4428"
/notes="Residues 1 to 289 of 289 are 100 pct identical to
residues 1 to 289 of a 289 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_291009.1"
/codon_start=1
/transl_table=1
/product="right origin-binding protein"
/protein_id="AAN45841.1"
/db_xref="GI:24054964"
/translation="MDAGITRDLLIMLEGLDQPLSLDNVAAKAGYSKWLQRMFKD
VTGHAGAYIRARLSKSAVALRLTARPIIDIALQYFDSQQFTFAFKQFOPTAL
YRSPREMAFGIRPLRLGSEFTMBEHFVLLDPLVPGDSQFTFAFKQFOPTAL
YOFHDFLGNAPTLPVLYGNLTPSPQDDBOEVYTTALADODADGVUTGVHVM
LQGGSEVMEYEGELGTGVEERILTVYGTCHPMNLIRKKQDIERTPADAKAGRP
INRCELLIPRR"
6797..7270
/gene="creA"
CDS
```

```

/locus_tag="SF4429"
6797..7270
/gene="creA"
/locus_tag="SF4429"
/notes="Residues 1 to 157 of 157 are 100 pct identical to
residues 1 to 157 of a 157 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_291010.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAN45842.1"
/db_xref="GI:24054965"
/translation="MKYKHLISLSLIMGLPLAHERIGSDYTPKAIQDPDKTIVEA
PDDPVKNVTCVSRKATGKIGKIGLAEPDSDAISCCQVGP1ELSDRIKNGAOGF
VYFKRSTLVFSPSLQVAFPYAKRNLALYLAISDKVVGSPKAIASVPMVPMRG"
7283..7972
/gene="creB"
/locus_tag="SF4430"
7283..7972
/gene="creB"
/locus_tag="SF4430"
/notes="Residues 1 to 229 of 229 are 98 pct identical to
residues 1 to 229 of a 229 aa protein from Escherichia
coli K12 ref: NP_418615.1"
/codon_start=1
/transl_table=1
/product="catabolic regulation response regulator"
/protein_id="AAN45843.1"
/db_xref="GI:24054966"
/translation="MORETVLIVDEDOGIADTLVYMLQOEGRVAVEFERGLPYLDAR
AQVDDWMLIDVGLPDISGFRLCQQLALHPLVPLTNRSEBVLRLLEIGADYV
AKQSPREVCARVATLIRKVRKFTSPVIRIGHFELNPEAAQISWFDPLITRRE
LLAKTLKSGFRWSSRQQLMDSVWEDADOTYRTVDYIHKTRAKRLAINPLSPINT
HRGMYSLRGL"
7972..9396
/gene="creC"
/locus_tag="SF4431"
7972..9396
/gene="creC"
Query Match 98.2%; Score 910.2; DB 1; Length 11661;
Best Local Similarity 99.6%; Pred. No. 1.9e-259;
Matches 923; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 GTCATGTTACGCCGATCATGTTAATTTGACGATGATTCAGGACGACTTTTGT 60
Db 11681 GTCATGTTACGCCGATCATGTTAATTTGACGATGATTCAGGACGACTTTTGT 11622
QY 61 ACTTCTGTTTGCATTTAGTTAGTGGCAATTTAGGTAGCAACATGACACCGGCACATTTCT 120
Db 11621 ACTTCTGTTTGCATTTAGTTAGTGGCAATTTAGGTAGCAACATGACACCGGCACATTTCT 11562
QY 121 TATCGTTGAAGCGAGTTGTTAAGCGCAACACGTTGAAAGTATTTTTCGAACGGGAAG 180
Db 11561 TATCGTTGAAGCGAGTTGTTAAGCGCAACACGTTGAAAGTATTTTTCGAACGGGAAG 11502
QY 181 CTATGATGTTTTCGAAGCAGATGGCGCGGAATCATCAGATGCTCTGTAATTTGA 240
Db 11501 CTATGATGTTTTCGAAGCAGATGGCGCGGAATCATCAGATGCTCTGTAATTTGA 11442
QY 241 CATCAACCTGGTATCATGATATCATCTCCGGGTAAAGACGGTCTTGTAGCGCG 300
Db 11441 CATCAACCTGGTATCATGATATCATCTCCGGGTAAAGACGGTCTTGTAGCGCG 11382
QY 301 TGAACCTGGCAGAGCGAATGTTGCGTTGATGTTCTGACCTGGCGCGGAACGAAGT 360
Db 11381 TGAACCTGGCAGAGCGAATGTTGCGTTGATGTTCTGACCTGGCGCGGAACGAAGT 11322
QY 361 CGATAAATTCCTGGCCTCGAATTCGTTGAGATGATCATCATCAACCAACCGTTCAACC 420
Db 11321 CGATAAATTCCTGGCCTCGAATTCGTTGAGATGATCATCATCAACCAACCGTTCAACC 11262
QY 421 GCGTGAACGATTCGTGACGACGACGACGATCTGCTCCGTACCATGATCTGGTACTGT 480
```

Db 11261 GCGTGAACGACGATTCGTGCAAGCAACGCTGCTCCCGACATGATCTGGGTACTGT 11202
OY 481 CAGCGAAGACGCTGCTAGCGTTGAAAGCTACAGTTCAATGTTGGAACTGGACATCA 540
Db 11201 CAGCGAAGACGCTGCTAGCGTTGAAAGCTACAGTTCAATGTTGGAACTGGACATCA 11142
OY 541 CAGCGCTTGGTATGCGCCCTGATGCGAGCACTACAAAGCTCCGCGCAGCGATTCCG 600
Db 11141 CAGCGCTTGGTATGCGCCCTGATGCGAGCACTACAAAGCTCCGCGCAGCGATTCCG 11082
OY 601 CCGCATGCTTCACTTCTGTGAAAAACCGAGCAAAATTCAGTCCCGCTGTAATCTGCTGA 660
Db 11081 CCGCATGCTTCACTTCTGTGAAAAACCGAGCAAAATTCAGTCTGCTGTAATCTGCTGA 11022
OY 661 GAAATGACCGCGCGCTGAGCTGAAACCGCAGCAGCCGACTGTAGAGTGCAGATCCGCGC 720
Db 11021 GAAATGACCGCGCGCTGAGCTGAAACCGCAGCAGCCGACTGTAGAGTGCAGATCCGCGC 10962
OY 721 TATTCGTAAACATTTTGAATCTACGCCGATACGCCGGAATCATGCCACATTCAACG 780
Db 10961 TATTCGTAAACATTTTGAATCTACGCCGATACGCCGGAATCATGCCACATTCAACG 10902
OY 781 TGAAGTTATCGCTTCTGCGGTATCTGGAAGTTATCGCTTTACACCTGCAAAAA 840
Db 10901 TGAAGTTATCGCTTCTGCGGTATCTGGAAGTTATCGCTTTACACCTGCAAAAA 10843
OY 841 AACGGGCTTTAGGCGCGTTTATTTTCAACCTTATTCAGATTAATTCAGTAACTCATC 900
Db 10842 AACGGGCTTTAGGCGCGTTTATTTTCAACCTTATTCAGATTAATTCAGTAACTCATC 10783
OY 901 GTCCGTTGTAACCTCTTTACTGCGCTT 927
Db 10782 GTCCGTTGTAACCTCTTTACTGCGCTT 10756

RESULT 9
AE016993/c
LOCUS Shigella flexneri 2a str. 2457T DNA linear BCT 22-APR-2003
DEFINITION genome.
ACCESSION AE016993
VERSION AE016993.1 GI:30043695
KEYWORDS
SOURCE Shigella flexneri 2a str. 2457T
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 225944)
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
JOURNAL 12704152
PUBMED 2 (bases 1 to 225944)
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..225944
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement(271..867)
gene

CDS
/gene="fimb"
/locus_tag="S4466"
complement(271..867)
/gene="fimb"
/locus_tag="S4466"
/function="regulator; Surface structures"
/note="residues 1 to 198 of 198 are 100.00 pct identical
to residues 1 to 198 of 198 from Escherichia coli K-12 :
B4313; involved in phase variation"
/codon_start=1
/transl_table=11
/product="recombinase; regulator for fimA"
/protein_id="AAP19415.1"
/db_xref="GI:30043696"
/translation="MSKRRYLTCKEVQAMQAVCGATGARDYCLILAYRHGRARISE
LIDLIHYODDLNENGRINIRLKNQFSTVPLRPDEEAVERTPOEARNMGADRTAI
RTSRGRSLRROAAVRIITDAGTACTGOTPHMRHCGYELARQADTRLIQYLV
GHRNIRHTVRYTASNAARFAGIMERNLINBKIKREV"
complement(1462..1947)
/gene="fimb"
/locus_tag="S4467"
complement(1462..1947)
/gene="fimb"
/locus_tag="S4467"
/function="regulator; Surface structures"
/note="residues 1 to 160 of 199 are 97.50 pct identical to
residues 1 to 160 of 200 from Escherichia coli K-12 :
B4312; involved in phase variation"
/codon_start=1
/transl_table=11
/product="recombinase; regulator for fimA"
/protein_id="AAP19416.1"
/db_xref="GI:30043697"
/translation="MRNKADNKKRNFITSEISLLKAANTGPATRYNCLILCFIH
GRASRICKRLRSIDILKAKCIYIRHLKNGPSTTHRLAKVQALKNWLSISPSVSHA
ESWVFLSRKNPLSRQPFHIIISGNAGLSELRHPMLRISCFALNMGIDTRL
I"
3569..4117
/gene="yJha"
/locus_tag="S4468"
3569..4117
/gene="yJha"
/locus_tag="S4468"
/note="residues 1 to 182 of 182 are 100.00 pct identical
to residues 60 to 241 of 241 from Escherichia coli K-12 :
B4311"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19417.1"
/db_xref="GI:30043698"
/translation="MESNTWTHIHKKENALINDOVEVNYAIKLDIDQWTVRPGMLT
HSSNSTRGYPLKLSWDATKDLNFRIRRYDKAYRQODLSGDSRDNVHRDGYVT
YHNSDFTPAWQTTLYSKONDYRANHKWATENAPVLQYHMTPTDTPYIEYDLRO
GVYNGRDLNSENSYRIGVSFKL"
4029..5243
/gene="yJht"
/locus_tag="S4469"
4029..5243
/gene="yJht"
/locus_tag="S4469"
/note="residues 1 to 404 of 404 are 98.76 pct identical to
residues 1 to 404 of 404 from Escherichia coli K-12 :
B4310"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19418.1"
/db_xref="GI:30043699"
/translation="MTTLTARVETTAETIYRKTVTALVCHLNCSPQETVYTNKTMAL
ATMNASFPAASVLPETVPVPSGTAINDTVYIGLSAGATWYKLDLPQADKKRTA
LAAFPGRQATISAFIDNLVYFGIGNSGLTQVFNVDHKNPKNTSKWSLASHA
PMGMAGHVTVFVHNGKAVYTVGVNOMIFNGYFEDLNEAGDSTALDKINAHYEDKQAD

/translation: "MKVITVTVPGKQRLAKKCPGSGADVDFOTFSGAIVNNSQKKVLL
 VMSVITQCHTDYNNVLSLSPKIVSKNRIKGRAMADAVTQFVAYDNLILKRNNA
 ELSGQDPHQLQHEQWVLSIASHKQVSNVPISSPQVSLILTKQGLLSAEIRRIE
 TQGLISGVETLQKSRKLNKAKIKDLOLSVETTLVRRIRIVLTR"
 1451. .2362
 /gene="STY4859"
 1451. .2362
 /note="N-termiun is similar to Salmonella typhimurium
 ITCALICA TR:09RPB6 (EMBL:AF164435) (74 aa) faeta scores:
 E(): 3e-27, 98.4% id in 74 aa, indicated by promoterless
 reporter gene insertion to be induced upon bacterial
 invasion of phagocytic and nonphagocytic cells (but having
 no affect on virulence). The adjacent 70 aa is highly
 similar to the neighbouring CDS Salmonella typhimurium
 hypochetrical protein TR:09RP5 (EMBL:AF164435) (70 aa)
 faeta scores: E(): 1.9e-20, 91.4% id in 70 aa"
 /codon start=1
 /transl table=11
 /product="hypochetrical protein"
 /protein_id="CAD003348.1"
 /db_xref="GI:16505620"
 /db_xref="TrEMBL:O820Y4"
 /translation: "MHTPPRYLSDLDYEDIPLEIEQIILLIINTLGNCSFYDM
 TVVQNSHSDYERATVICTLKEKGLFTLVASQVLEMSNLIKTDKDKKIIDPT
 ELAPNVITDITVLTGYETTLGEPDKALSELNIFHEHSIKYENGRIVRLRIRIKR
 ILCYLVSPEGYGEVDELDLAEKRLNDFACVVECHAVRSVTLRIHEKSLVHS
 QDSLIVTLNMLNARQVISAAYHKFPLIDVLESEIPLPMAPEKSHHNAWLIE
 LACSVIYLTERRSIFLQTORAEEDGVAP"
 2573. .3034
 /gene="STY4860"
 2573. .3034
 /note="S4imilar to Neisseria meningitidis putative
 periplasmic protein NMA0093 TR:CA884263 (EMBL:AL162754)
 (159 aa) faeta scores: E(): 1.2e-12, 34.4% id in 160 aa
 and Neisseria meningitidis hypochetrical protein MNM0783
 TR:AAFP1196 (EMBL:AE002432) (159 aa) faeta scores: E():
 1.3e-12, 34.4% id in 160 aa. Also similar to Haemophilus
 ducreyi hypochetrical protein TR:AAFP3378 (EMBL:AF219260)
 (149 aa) faeta scores: E(): 9.2e-11, 31.5% id in 146 aa
 and Treponema pallidum hypochetrical protein TP0895
 TR:083865 (EMBL:AE001259) (168 aa) faeta scores: E():
 6.9e-09, 33.6% id in 134 aa. Contains a probable
 N-terminal signal sequence"
 /codon start=1
 /transl table=11
 /product="putative exported protein"
 /protein_id="CAD003349.1"
 /db_xref="GI:16505621"
 /db_xref="TrEMBL:O820Y3"
 /translation: "MSTIILKPLFASLTLILRPFGTEGATLVKTVSAVYLDOT
 RARLMAVADHGVLTGSEDFDRAAONVNLKMPPTVFPNGPGKGVLAHPELAL
 DLRPRVLISQAGRCRLVSYHNAETLQKRGDADAIQALKLEQVKSLSH"
 complement (3051. .3716)
 /gene="STY4861"
 complement (3051. .3716)
 /note="S4imilar to Bacillus subtilis hypochetrical protein
 YCZH YCZH SM:YCZH_BACSU (OJ1482) (185 aa) faeta scores:
 E(): 7.4e-16, 31.4% id in 159 aa"
 /codon start=1
 /transl table=11
 /product="hypochetrical protein"
 /protein_id="CAD003350.1"
 /db_xref="GI:16505622"
 /db_xref="TrEMBL:O820Y2"
 /translation: "MKIDTVFAPSPGGSNASSLSRHRRAAILIIEHYGINAHRIOR
 AGHMMRGEDIVITLPHNHPFRPEQEEASVHFSANVGEPAVVVTLIHLDAQY
 STILVQVYSGALTAALASGLCGVGVCSYSPRIQVYHLALPCLALVITARPEAF
 DPRAAMOTLEKRPVCCQKDYDARHGFCDADATFPALASHOYMDVSAFIRDTICANT
 SPD"

misc_feature complement(3375..3404)
/gene="STY4861"
/note="PS00120 Lipases"
4066..4839
/gene="uxuR"
/note="synonym: STY4862"
4066..4839
/gene="uxuR"
/note="Similar to Escherichia coli uxu operon
transcriptional regulator uxuR SW:UXUR_ECOLI (P39161) (257
aa) fasta scores: E(): 0, 89.9% id in 257 aa, and to
Erwinia chrysanthemi exu regulon transcriptional regulator
exuR SW:EXUR_ERWCH (Q9X9E0) (259 aa) fasta scores: E(): 0,
residues 34 to 55, score 975(+2.51 SD).
Fasta hit to PDHR_ECOLI (254 aa), 32% identity in 221 aa
overlap
Fasta hit to EXUR_ECOLI (258 aa), 48% identity in 255 aa
overlap
Orthologue of E. coli uxuR (UXUR_ECOLI); Fasta hit to
UXUR_ECOLI (257 aa), 90% identity in 257 aa overlap"
/codon_start=1
/transl_table=1
/product="uxu operon transcriptional regulator"
/protein_id="CAD03351.1"
/db_xref="GI:16505623"
/db_xref="GOA:Q820Y1"
/db_xref="TrEMBL:Q820Y1"
/translation="MKSNTSQQRPYQEVGAMIRDLIVQPPRGERLPERRIARLN
VTRIVAEALIMELIKGLVEVRGAGIYVDSADNNMBCADYVHNCNDGAPPELLQAR
QLLESNTAEPAALQATREDIIRKQALQLEERLASARGSSGMORHQAIAEKTH
NSMLVLFROSQWQMRNNPMWLDHSHLGDTTYRKMLGDHKOILALIKKQARAKL
AMNQHLENYQRILERSNVDDITFDGILFESWPLDVDA"
4105..4284
/gene="uxuR"
/note="Pfam match to entry PF00392 gntR, Bacterial
regulatory proteins, gntR family, score 107.20, E-value
1.4e-31"
4162..4236
/gene="uxuR"
/note="PS00043 Bacterial regulatory proteins, gntR family
signature"
complement(4840..5853)
/gene="trpS2"
/note="synonym: STY4863"
complement(4840..5853)
/gene="trpS2"
/EC_number="6.1.1.2"
/note="Similar to Clostridium longisporum
tryptophanyl-tRNA synthetase Trps or Trsa SW:SYW_CLOLO
(Q46127) (341 aa) fasta scores: E(): 0, 54.8% id in 330
aa, and to Thermotoga maritima tryptophanyl-tRNA
synthetase Trps SW:SYW_THEMA (Q9WY2) (328 aa) fasta
scores: E(): 0, 36.5% id in 329 aa. Note that the E.coli
trps ortholog is STY4315"
/codon_start=1
/transl_table=1
/product="probable tryptophanyl-tRNA synthetase"
/protein_id="CAD03352.1"
/db_xref="GI:16505624"
/db_xref="GOA:Q820Y0"
/db_xref="TrEMBL:Q820Y0"
/translation="WNTLPILITGDRPTGSLHGHVSLRQVALQODHOQVLIAD
LOGLTGNSNPQIKRDNIPQVADYLAVIDPALTTICQSLAPLALATLQVNIAT
TVARERNPTVKNEIAQKGTETSLPYGFAFYISQAADITAKAEVNPVGDQIWMIO
TNEIVHKNSLFPSSPYLPPCCALSDTGRLEFIDSAVMSLVNTLLLSGSETHIR
AVSAMTYTPNHKISDPGKIRSNVFTWLDLPHDKAVAMKHYQGGIGDVVCKN
ELFETQLGLIAPIRERRATFLADKMLMELIKGSERAHEVTQKTLQVKKGLGLPTL
FQV"
complement(5170..5835)
/gene="trpS2"
/note="Pfam match to entry PF00579 trna-synt_1b, trna

Query Match		77.2%;	Score 715.6;	DB 1;	Length 88037;
Best Local Similarity		92.2%;	Pred. No. 2.9e-201;		
Matches 754;		Conservative 0;	Mismatches 64;	Indels 0;	Gaps 0;
Qy	1	GTGATGTTACGGCGATCATGTTAATTTGACAGATGATGACGAGCTGAGCACTTTTGT	60		
Db	86791	GTGATGTTACGGCGATCATGTTAATTTGACAGATGATGACGAGCTGAGCACTTTTGT	86732		
Qy	61	ACTTCCTGTTTGCATTTAGTGGCAATTTAGGTAGCAACATGACAGCCCGCACTTCT	120		
Db	86731	ACTTCCTGTTTGCATTTAGTGGCAATTTAGGTAGCAACATGACAGCCCGCACTTCT	86672		
Qy	121	TATCGTTGAAGCGAGTTGGTAAACGCAACAGCTTGAAGAAATTTTGAAGCGAAGC	180		
Db	86671	TATCGTTGAAGCGAGTTGGTAAACGCAACAGCTTGAAGAAATTTTGAAGCGAAGC	86612		
Qy	181	CTATGATGTTTTCGAAGCAGACAGATGAGCGCGGAATGCAATCTCTCGAATATGA	240		
Db	86611	CTATGATGTTTTCGAAGCAGACAGATGAGCGCGGAATGCAATCTCTCGAATATGA	86552		
Qy	241	CATCAACCTGGTGATGATGATATGATATGATCTGCGGGTAAAGACGGCTCTTGTAGCGCG	300		
Db	86551	CATCAACCTGGTGATGATGATATGATATGATCTGCGGGTAAAGACGGCTCTTGTAGCGCG	86492		
Qy	301	TGAATCGCGCAGCAGAGCGAATGTTGGGTGATGTTCTGACCTGGCCGCTGACAGAGT	360		
Db	86491	TGAATCGCGCAGCAGAGCGAATGTTGGGTGATGTTCTGACCTGGCCGCTGACAGAGT	86432		
Qy	361	CGATAAATTTCTGGGCTTCGAATTCGTGAGATGATACATCAACAAACGGTTCACACC	420		
Db	86431	CGACAAATTTCTGGGCTTCGAATTCGTGAGATGATACATCAACAAACGGTTCACACC	86372		
Qy	421	GGTGAACGACAGATTCGTGACAGCAACCTACTGTCCTCCGTAACATGCTGGTACTGT	480		
Db	86371	AGCGAATGACATTCGTGACAGCAACCTACTGTCCTCCGTAACATGCTGGTACTGT	86312		
Qy	481	CAGCGAAGACGTCGTGACGTTGAAAGCTACAACTTCAATGTTGGGAATGACATCAA	540		
Db	86311	CAGCGAAGACGTCGTGACGTTGAAAGCTACAACTTCAATGTTGGGAATGACATCAA	86252		
Qy	541	CAGCGCTTCGTTGATTCGGCCTTATGCGAGACGACGATCAAGCTGCGCGACGATTCG	600		
Db	86251	CAGCGCTTCGTTGATTCGGCCTTATGCGAGACGACGATCAAGCTGCGCGACGATTCG	86192		
Qy	601	CGCATGCTTCACTTCTGTGAAAACCGAGGAAATTCAGTCCCGCTGAACTGCTGAA	660		
Db	86191	CGCATGCTTCACTTCTGTGAAAACCGAGGAAATTCAGTCCCGCTGAACTGCTGAA	86132		
Qy	661	GAATAATGACCGGCGGTGAGCTGAAAACCGACAGACCGTACTGTAGACGTGACGATCCGCG	720		
Db	86131	GAATAATGACCGGCGGTGAGCTGAAAACCGACAGACCGTACTGTAGACGTGACGATCCGCG	86072		
Qy	721	TATTCGTAACATTTGCAATCTACCGCGGATACCGCGGAATATCATGCCACATTCACGG	780		
Db	86071	TATTCGTAACATTTGCAATCTACCGCGGATACCGCGGAATATCATGCCACATTCACGG	86012		
Qy	781	TGAAGTTATCGCTTCTGCGGTGATGCAAGATTAT	818		
Db	86011	TGAAGTTATCGCTTCTGCGGTGATGCAAGATTAT	85974		
RESULT 11					
AE016849/c					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Salmonella enterica subsp. enterica serovar Typhi Ty2					
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;					
Enterobacteriaceae; Salmonella.					

```
REFERENCE      1 (baees 1 to 284233)
AUTHORS        Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
                Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
TITLE          Comparative Genomics of Salmonella enterica Serovar Typhi Strains
                Ty2 and CT18
JOURNAL        J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE        2 (baees 1 to 284233)
PUBMED        1264504
                22531367
AUTHORS        Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
                Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
TITLE          Direct Submission
JOURNAL        Submitted (25-SEP-2002) Laboratory of Genetics, University of
                Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES       source
                1. 284233
                /organism="Salmonella enterica subsp. enterica serovar
                Typhi Ty2"
                /mol_type="genomic DNA"
                /strain="Ty2"
                /serovar="Typhi"
                /sub_species="enterica"
                /db_xref="taxon:209261"
                301. .489
                /locus_tag="t4354"
                /locus_tag="t4354"
                /note="corresponds to STY4663 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071807.1"
                /db_xref="GI:29140244"
                /translation="MTGKPSERTGRTIGSEMMVRDCCFGNEYLHAGEAFVSENHDA
                WVVGDPVCVALDTHFLR"
                614. .2125
                /locus_tag="t4355"
                614. .2125
                /locus_tag="t4355"
                /note="corresponds to STY4664 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative DNA helicase"
                /protein_id="AA071808.1"
                /db_xref="GI:29140245"
                /product="t4355"
                /note="corresponds to STY4665 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071809.1"
                /db_xref="GI:29140246"
                /translation="MRKYKSWMEKKNGPROIEBNSGNKKRLIPTCTDGYFNPATISD
                LHOHEROOLRLIMENSAIPKRPFRYFLPLHGVVGGCORLPASAGKPAVTDGAV
                DVIOCTTVFAVRLSKGYMLPRGASAEQSNQSAVAYYYAALPHSLGRWNTIBEL
                RGAIVRPGISVPEEPYRFRKAPDIAQOYVIGELIAYLAFPEPVLQWLGKPDILR
                TLAFISGRVSDAMDITDIIVNEIVHAGGIPLGFLIAGEQTSLEAEILINPAVAVPV
                LKGSSTTPPDVAVVWDGQETTVLLISLSDAKPBRGCGPTAETDPVLQVMSITLG
                GLTNTDTSQPEKEGVNTEITDWNAGSFFFEVAVALETSAAEBEIVNTVTTECTSSSGG
                ISLSSTELGDLFWMSWLRDGLREGDIPVNTADACVHLTCGFVPI SVPGFELKSHS
                RSCSSGLKSGSKRGKEQVQAPEKRRKRVSDSRFMWCCLYEEPGRGYKGLTGYYLK
                MSEIYANGNPPDLSLFLKVIN"
                3662. .4875
                /locus_tag="t4357"
                3662. .4875
                /locus_tag="t4357"
                /note="corresponds to STY4666 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="probable phage integrase"
                /protein_id="AA071810.1"
                /db_xref="GI:29140247"
                /translation="METNTTWOOLIDEYFPAPKPLRSASEMSYTKVFPVWGPLSC
                PNDVYTKYLARRRPLTKKKSIGRTNNKVAHMRALPNYVGIORGILHLDENFPNNSV
                VEPDKRKRTLTQAOQTEIYAVIQIMEQYENENETGLKYSICALPAMFWLTVLDTLY
                TGIRONQHLIRLNDVLDREGOIRLITTECKKHKEYVVISPLRPLTCLMEKAOSE
                GLKGNDRLFNIALFTGKDPALIGDMDSPVRAFFRRLSECCFALISPHRFRHLATEM
                MKMPEQNLHMAQSVLGSNNKSTLEVENDIAMVGRALBAQFMQIKAAVASIYSGLT
                KRR"
                5302. .5595
                /locus_tag="t4358"
                5302. .5595
                /locus_tag="t4358"
                /note="corresponds to STY4667 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071811.1"
                /db_xref="GI:29140248"
                /translation="MPVANSMAKRETTNLRIKPARERDLIDRAAKARGKNCDFVLEA
                AAAAEALIEORITIMADPEAYQAFILRLOAPAPDALRKTMTQTPAPMBQER"
                5592. .6080
                /locus_tag="t4359"
                5592. .6080
                /locus_tag="t4359"
                /note="corresponds to STY4668 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AA071812.1"
                /db_xref="GI:29140249"
                /translation="MISAPEPLHAGHILTPCCGDSMDNMLKORAMKQVTCASRTF
                VCCGSDVLAAYSLASSAVTNTAPGRRRAMPDPIPVVILGRILAVDRSLHGQGVGR
                ALVRDAGLRVIOVAETIGIRGMLVHALSDBARFYLRVGFEPSPMDPMLMVLGDLV
                HA"
                complement (6260. .6712)
                /locus_tag="t4360"
                complement (6260. .6712)
                /locus_tag="t4360"
                /note="corresponds to STY4669 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative membrane protein"
                /protein_id="AA071813.1"
                /db_xref="GI:29140250"
                /translation="MEQCKNDAILHEIKYNSKHIDFERSQANSOGIWLFI STLGCWSV
                NPLIOVATAIILLFCIFPNSKODMTEKRAFKRIEDIAKDDISNLIGSRKARLYDL
                GLVEKTKRAIKVLTSPFIYICYIFYSISFLVFSNLPFRKRLIFN"
                complement (7349. .9634)
                /locus_tag="t4361"
                complement (7349. .9634)
                /locus_tag="t4361"
                /note="corresponds to STY4670 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
```

QY	361	CGATTAATTTCTGGCCCTCGAAATCGGTGATGATGACTACATCAACCAACCGTTCAACC	420		
Db	282627	CGACAAAAATTCGGCCCTCGAAATCGGTGATGATGACTACATCAACCAACCGTTAAATCC	282566		
QY	421	GGTGAATCGACGATTTGGTGACCGCAACCTACTCTCCCGTACATGATTTGGTACTGT	480		
Db	282567	ACGGGAATTGACATCCGTCCCGCTTAACCTGCTGTCCCGTACATGATTTGGTACTGT	282508		
QY	481	CAGCGAAGAACGTGCTGATGCTGTAAGCTTCAAGTTCAATGCTTGGGAACTGGACATCAA	540		
Db	282507	CAGCGAAGAACGTGCTGATGCTGTAAGCTTCAAGTTCAATGCTTGGGAACTGGATATCAA	282448		
QY	541	CAGCGTTCGTTGATTCGCGCCCTGATGCGACGATCAAGCTCCCGCGACGATTCGG	600		
Db	282447	CAGCGTTCGTTGATTCGCGCCCTGATGCGACGATCAAGCTCCCGCGACGATTCGG	282388		
QY	601	CGCATATGCTTCACTTCTGTGTAATAACCGACGCAAAATTCAGTCCCGTGTGAATCTGTGA	660		
Db	282387	CGCATATGCTTCACTTCTGTGTAATAACCGACGCAAAATTCAGTCCCGTGTGTGA	282328		
QY	661	GAATATGACCGCGCGGTGAGTGAACCGGACGACCGCTACTGTGACGTGACATCCGCCG	720		
Db	282327	AAAAATGACCGGACCGGAGTGAACCGGACGACCGCTACTGTGACGTGACATCCGCCG	282268		
QY	721	TATTCGTAACATTTGGAATCTACCGCGGATACCGCGGAATATATGCGCACCATTCACGG	780		
Db	282267	TATTCGTAACATTTGGAATCTACCGCGGATACCGCGGAATATATGCGCACCATTCACGG	282208		
QY	781	TGAAGTTATCGCTTCTGCGGTGATCTGGAATTAAT	818		
Db	282207	CGAAGTTATCGCTTCTGCGGTGATCTGGAATTAAT	282170		
RESULT 12	AE008916/c	13852 bp	DNA	linear	BCT 23-APR-2003
LOCUS	AE008916	13852 bp	DNA	linear	BCT 23-APR-2003
DEFINITION	Salmonella typhimurium LT2, section 220 of 220 of the complete genome.				
ACCESSION	AE008916	AE006468			
VERSION	AE008916.1	GI:16423153			
KEYWORDS					
SOURCE					
ORGANISM	Salmonella typhimurium LT2				
REFERENCE	Salmonella typhimurium LT2				
AUTHORS	Bacteroides; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.				
	1 (bases 1 to 13852)				
	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,				
	Latreille, P., Courtney, L., Portolillo, S., All, J., Dante, M., Du, F.,				
	Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,				
	Graw, N., Mulvane, E., Ryan, E., Sun, H., Florea, L., Miller, W.,				
	Stoneman, T., Nhan, M., Waterston, R., and Wilson, R.K.				
	Complete genome sequence of Salmonella enterica serovar Typhimurium				
	LT2				
	Nature 413 (6858), 852-856 (2001)				
JOURNAL	21534948				
MEDLINE	11677609				
PUBMED	2 (bases 1 to 13852)				
REFERENCE					
AUTHORS					
CONSTRM					
TITLE					
JOURNAL					
COMMENT					

The Salmonella typhimurium Genome Sequencing Project
 Direct Submission
 Submitted (29-MAR-2001) Genome Sequencing Center, Department of
 Genetics, Washington University School of Medicine, 4444 Forest
 Park Boulevard, St. Louis, MO 63108, USA
 COMMENT Supported by NIH grant 5U01 AI43283

Coding sequences below are predicted from manually evaluated
 computer analysis, using similarity information and the programs;
 GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
 GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
 Encyclopedia of Genes and Genomes; <http://www.genome.jp/kegg/>

<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and Reguondb:
http://kinich.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES

source

Location/Qualifiers

1..13852

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

90..640

/gene="crea"

/note="synonym: STM4587"

90..98

/gene="crea"

/note="putative -35 signal for crea";

RegulonDB:STMLTH004682"

117..125

/gene="crea"

/note="putative -10 signal for crea";

RegulonDB:STMLTH004682"

156..161

/gene="crea"

/note="putative RBS for crea; RegulonDB:STMSIH004447"

167..640

/gene="crea"

/note="similar to E. coli orf, hypothetical protein (AAC77350.1); Blastp hit to AAC77350.1 (157 aa), 89%

identity in aa 1 - 157"

/codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="AAL23402.1"

/db_xref="GI:16423154"

/translation="MKKSLVLESLILLGQSRARQISVDVVFQKFGPDHKIVYEA FDDPVKAVTCYSRAKGTGIGGLAEDTDAISCCQVGEIELSDKIKNGKAGE VVFKKRTSLIFPSLQVRFYDEKRTLAYLAISDKVVDSPKNAISAVPMWRQ"

641..1342

/gene="creb"

/note="synonym: STM4588"

641..646

/gene="creb"

/note="putative RBS for creb; RegulonDB:STMSIH004448"

653..1342

/gene="creb"

/note="OmpR family; similar to E. coli catabolic regulation response regulator (AAC77351.1); Blastp hit to AAC77351.1 (229 aa), 83% identity in aa 1 - 229"

/codon_start=1

/transl_table=11

/product="response regulator in two-component regulatory system with CreC"

/protein_id="AAL23403.1"

/db_xref="GI:16423155"

/translation="MQQPVWLVEDEQGIADTLIYLQEGFTVELFARGPALAEKAR QORPDVILDVGLPDISGELCQLEERHPLILFLTARSDEVDRLGLGELGADYV AKPSREVSARVRLRLRVKPKAAPSPPVRTGFELNEPDAQIAWGTPLSLTRYEF LLKTLILSPERYSRQQLMDIWDSDAQEFDDTVTHIKTLAKLRALNPELSPINT HSGMGSLSVS"

1130..2766

/gene="crec"

/note="synonym: STM4589"

RBS

1330..1335

/gene="creb"

/note="putative RBS for crec; RegulonDB:STMSIH004449"

1342..2766

/gene="crec"

/EC_number="2.7.3.-"

/note="senses catabolite repression; similar to E. coli catabolite repression sensor kinase for Phob; alternative sensor for pho regulon (AAC77352.1); Blastp hit to AAC77352.1 (474 aa), 80% identity in aa 1 - 474"

/codon_start=1

/transl_table=11

/product="sensory kinase (alternative) in two-component regulatory system with Creb (or alternatively Phob)"

/protein_id="AAL23404.1"

/db_xref="GI:16423156"

/translation="MRIGRLLGVELYAVVAAPVLSFNOEIKRVRATEGLID TITLALARDLISGNPTDGLAKAFADQIRPRANISGIVKRNETHIYTDAG KVLFDSENKAVQDYSRMNDVWLTIRGQASSTADADPSTVWYVAPITISDRI IGLTVGKENAMAPVIRKSERRIILASAVILGIALGAGVWVWVNRISARLTRAD SVENRPLPALPAGSSELRKLAQLESMRIKLEGRKEIQYVYALTHEKSLPLAIVG AAEILRECPADVVRFTENIILQNTRMQALVETLIRORLENRODIPLAPVDELFTOLSEKRSQLAKKITTLTRPSSLVVADALDAQNGVUDNATDPENGVITLS AQPWGEKALIQVTDGCGIPDFALPEIFDFRYSLPRENGKSSGGLAFVSEARLLN GEVTLGNREGEGLASLTLRHPT"

2812..4173

/gene="cred"

/note="synonym: STM4590"

2812..2817

/gene="cred"

/note="putative RBS for cred; RegulonDB:STMSIH004450"

2824..4173

/gene="cred"

/note="similar to E. coli tolerance to colicin E2 (AAC77353.1); Blastp hit to AAC77353.1 (450 aa), 73%

identity in aa 1 - 450"

/codon_start=1

/transl_table=11

/product="tolerance to colicin E2 protein"

/protein_id="AAL23405.1"

/db_xref="GI:16423157"

/translation="MLKSPILMKMTITLGGAMTLLILPLMMVHTTVERADYSHVEA IKQSTSPQKVVGPVIAIPVTELYTLVEEKEVQKRSYLFWLPESLVEGNQVNEA RIGTIYQGVMTDPAIKAEFDVARLHELRNITLTKPFIIVGVGDARGLSAVKAPO VNGEITLVPGTGLPESRGGIHLPLPDSQMATRNLTLDKSLNLSGGSFSLVPGRSS EKTLSNMPHPNPNVDFPLGKREISGSGQAGQWOTSRFATNIGEPADAOXKDMDLPS AFSVAVSTPADDYQUTDRATKRAIILITITTPAFVFFETLTCRLHPMYTLYVGLSLV MRYLLALASERHIGFTPAWIASLVGALMNSVYLQAVLQGMNSVLPTVALALADGVW MGLRSDESSLIGTGVLLALAGVWFLTRHLDWYLSQCKRKSLLPPVXDDELRLMK"

complement (4231..5324)

/gene="sthe"

/note="synonym: STM4591"

complement (4231..5316)

/gene="sthe"

/codon_start=1

/transl_table=11

/product="putative major fimbrial subunit"

/protein_id="AAL23406.1"

/db_xref="GI:16423158"

/translation="MRRLYLALILFLAVSGHSYASCRRSGNEGATITTPPSQLVVDH AYTAGSVLQSGVSTSEVTMDGCSRDYVGVGLYEGSAQSTSAIINANDGNPFF STGISGVGAIKTQINAGPYDNVMPIDNTYHNGDKTKHAAPAVNVLVAGPIT SGTATGPOSPLARVFRDSDATBEDSGDILTHLVLGNTQMLKAMCGCVERPATVLDGS VNLGFSANSQTRGTEBODILTCBOGTASALSQRPASGNNPDNSVITQLSNASAPTS ATGVGVQIGIQAPDAGFTGSLPINOKIDLFTHTITTNADGQTVNGGTMNSTITLKI SARRYKTAATVAGQANATATLNTYN"

complement (5319..5324)

/gene="sthe"

/note="putative RBS for sthe; RegulonDB:STMSIH004451"

complement (5357..5929)

/gene="sthd"

/note="synonym: STM4592"

complement (5357..5914)

RBS

1330..1335

/gene="creb"

/note="putative RBS for crec; RegulonDB:STMSIH004449"

1342..2766

/gene="crec"

/EC_number="2.7.3.-"

/note="senses catabolite repression; similar to E. coli catabolite repression sensor kinase for Phob; alternative sensor for pho regulon (AAC77352.1); Blastp hit to AAC77352.1 (474 aa), 80% identity in aa 1 - 474"

/codon_start=1

/transl_table=11

/product="sensory kinase (alternative) in two-component regulatory system with Creb (or alternatively Phob)"

/protein_id="AAL23404.1"

/db_xref="GI:16423156"

/translation="MRIGRLLGVELYAVVAAPVLSFNOEIKRVRATEGLID TITLALARDLISGNPTDGLAKAFADQIRPRANISGIVKRNETHIYTDAG KVLFDSENKAVQDYSRMNDVWLTIRGQASSTADADPSTVWYVAPITISDRI IGLTVGKENAMAPVIRKSERRIILASAVILGIALGAGVWVWVNRISARLTRAD SVENRPLPALPAGSSELRKLAQLESMRIKLEGRKEIQYVYALTHEKSLPLAIVG AAEILRECPADVVRFTENIILQNTRMQALVETLIRORLENRODIPLAPVDELFTOLSEKRSQLAKKITTLTRPSSLVVADALDAQNGVUDNATDPENGVITLS AQPWGEKALIQVTDGCGIPDFALPEIFDFRYSLPRENGKSSGGLAFVSEARLLN GEVTLGNREGEGLASLTLRHPT"

2812..4173

/gene="cred"

/note="synonym: STM4590"

2812..2817

/gene="cred"

/note="putative RBS for cred; RegulonDB:STMSIH004450"

2824..4173

/gene="cred"

/note="similar to E. coli tolerance to colicin E2 (AAC77353.1); Blastp hit to AAC77353.1 (450 aa), 73%

identity in aa 1 - 450"

/codon_start=1

/transl_table=11

/product="tolerance to colicin E2 protein"

/protein_id="AAL23405.1"

/db_xref="GI:16423157"

/translation="MLKSPILMKMTITLGGAMTLLILPLMMVHTTVERADYSHVEA IKQSTSPQKVVGPVIAIPVTELYTLVEEKEVQKRSYLFWLPESLVEGNQVNEA RIGTIYQGVMTDPAIKAEFDVARLHELRNITLTKPFIIVGVGDARGLSAVKAPO VNGEITLVPGTGLPESRGGIHLPLPDSQMATRNLTLDKSLNLSGGSFSLVPGRSS EKTLSNMPHPNPNVDFPLGKREISGSGQAGQWOTSRFATNIGEPADAOXKDMDLPS AFSVAVSTPADDYQUTDRATKRAIILITITTPAFVFFETLTCRLHPMYTLYVGLSLV MRYLLALASERHIGFTPAWIASLVGALMNSVYLQAVLQGMNSVLPTVALALADGVW MGLRSDESSLIGTGVLLALAGVWFLTRHLDWYLSQCKRKSLLPPVXDDELRLMK"

complement (4231..5324)

/gene="sthe"

/note="synonym: STM4591"

complement (4231..5316)

/gene="sthe"

/codon_start=1

/transl_table=11

/product="putative major fimbrial subunit"

/protein_id="AAL23406.1"

/db_xref="GI:16423158"

/translation="MRRLYLALILFLAVSGHSYASCRRSGNEGATITTPPSQLVVDH AYTAGSVLQSGVSTSEVTMDGCSRDYVGVGLYEGSAQSTSAIINANDGNPFF STGISGVGAIKTQINAGPYDNVMPIDNTYHNGDKTKHAAPAVNVLVAGPIT SGTATGPOSPLARVFRDSDATBEDSGDILTHLVLGNTQMLKAMCGCVERPATVLDGS VNLGFSANSQTRGTEBODILTCBOGTASALSQRPASGNNPDNSVITQLSNASAPTS ATGVGVQIGIQAPDAGFTGSLPINOKIDLFTHTITTNADGQTVNGGTMNSTITLKI SARRYKTAATVAGQANATATLNTYN"

complement (5319..5324)

/gene="sthe"

/note="putative RBS for sthe; RegulonDB:STMSIH004451"

complement (5357..5929)

/gene="sthd"

/note="synonym: STM4592"

complement (5357..5914)

RBS

1330..1335

/gene="creb"

/note="putative RBS for crec; RegulonDB:STMSIH004449"

1342..2766

/gene="crec"

/EC_number="2.7.3.-"

/note="senses catabolite repression; similar to E. coli catabolite repression sensor kinase for Phob; alternative sensor for pho regulon (AAC77352.1); Blastp hit to AAC77352.1 (474 aa), 80% identity in aa 1 - 474"

/codon_start=1

/transl_table=11

/product="sensory kinase (alternative) in two-component regulatory system with Creb (or alternatively Phob)"

/protein_id="AAL23404.1"

/db_xref="GI:16423156"

/translation="MRIGRLLGVELYAVVAAPVLSFNOEIKRVRATEGLID TITLALARDLISGNPTDGLAKAFADQIRPRANISGIVKRNETHIYTDAG KVLFDSENKAVQDYSRMNDVWLTIRGQASSTADADPSTVWYVAPITISDRI IGLTVGKENAMAPVIRKSERRIILASAVILGIALGAGVWVWVNRISARLTRAD SVENRPLPALPAGSSELRKLAQLESMRIKLEGRKEIQYVYALTHEKSLPLAIVG AAEILRECPADVVRFTENIILQNTRMQALVETLIRORLENRODIPLAPVDELFTOLSEKRSQLAKKITTLTRPSSLVVADALDAQNGVUDNATDPENGVITLS AQPWGEKALIQVTDGCGIPDFALPEIFDFRYSLPRENGKSSGGLAFVSEARLLN GEVTLGNREGEGLASLTLRHPT"

2812..4173

/gene="cred"

/note="synonym: STM4590"

2812..2817

/gene="cred"

/note="putative RBS for cred; RegulonDB:STMSIH004450"

2824..4173

/gene="cred"

/note="similar to E. coli tolerance to colicin E2 (AAC77353.1); Blastp hit to AAC77353.1 (450 aa), 73%

identity in aa 1 - 450"

/codon_start=1

/transl_table=11

/product="tolerance to colicin E2 protein"

/protein_id="AAL23405.1"

/db_xref="GI:16423157"

/translation="MLKSPILMKMTITLGGAMTLLILPLMMVHTTVERADYSHVEA IKQSTSPQKVVGPVIAIPVTELYTLVEEKEVQKRSYLFWLPESLVEGNQVNEA RIGTIYQGVMTDPAIKAEFDVARLHELRNITLTKPFIIVGVGDARGLSAVKAPO VNGEITLVPGTGLPESRGGIHLPLPDSQMATRNLTLDKSLNLSGGSFSLVPGRSS EKTLSNMPHPNPNVDFPLGKREISGSGQAGQWOTSRFATNIGEPADAOXKDMDLPS AFSVAVSTPADDYQUTDRATKRAIILITITTPAFVFFETLTCRLHPMYTLYVGLSLV MRYLLALASERHIGFTPAWIASLVGALMNSVYLQAVLQGMNSVLPTVALALADGVW MGLRSDESSLIGTGVLLALAGVWFLTRHLDWYLSQCKRKSLLPPVXDDELRLMK"

complement (4231..5324)

/gene="sthe"

/note="synonym: STM4591"

complement (4231..5316)

/gene="sthe"

/codon_start=1

/transl_table=11

/product="putative major fimbrial subunit"

/protein_id="AAL23406.1"

/db_xref="GI:16423158"

/translation="MRRLYLALILFLAVSGHSYASCRRSGNEGATITTPPSQLVVDH AYTAGSVLQSGVSTSEVTMDGCSRDYVGVGLYEGSAQSTSAIINANDGNPFF STGISGVGAIKTQINAGPYDNVMPIDNTYHNGDKTKHAAPAVNVLVAGPIT SGTATGPOSPLARVFRDSDATBEDSGDILTHLVLGNTQMLKAMCGCVERPATVLDGS VNLGFSANSQTRGTEBODILTCBOGTASALSQRPASGNNPDNSVITQLSNASAPTS ATGVGVQIGIQAPDAGFTGSLPINOKIDLFTHTITTNADGQTVNGGTMNSTITLKI SARRYKTAATVAGQANATATLNTYN"

complement (5319..5324)

/gene="sthe"

/note="putative RBS for sthe; RegulonDB:STMSIH004451"

complement (5357..5929)

/gene="sthd"

/note="synonym: STM4592"

complement (5357..5914)

RBS

1330..1335

/gene="creb"

/note="putative RBS for crec; RegulonDB:STMSIH004449"

1342..2766

/gene="crec"

/EC_number="2.7.3.-"

/note="senses catabolite repression; similar to E. coli catabolite repression sensor kinase for Phob; alternative sensor for pho regulon (AAC77352.1); Blastp hit to AAC77352.1 (474 aa), 80% identity in aa 1 - 474"

/codon_start=1

/transl_table=11

/product="sensory kinase (alternative) in two-component regulatory system with Creb (or alternatively Phob)"

/protein_id="AAL23404.1"

/db_xref="GI:16423156"

/translation="MRIGRLLGVELYAVVAAPVLSFNOEIKRVRATEGLID TITLALARDLISGNPTDGLAKAFADQIRPRANISGIVKRNETHIYTDAG KVLFDSENKAVQDYSRMNDVWLTIRGQASSTADADPSTVWYVAPITISDRI IGLTVGKENAMAPVIRKSERRIILASAVILGIALGAGVWVWVNRISARLTRAD SVENRPLPALPAGSSELRKLAQLESMRIKLEGRKEIQYVYALTHEKSLPLAIVG AAEILRECPADVVRFTENIILQNTRMQALVETLIRORLENRODIPLAPVDELFTOLSEKRSQLAKKITTLTRPSSLVVADALDAQNGVUDNATDPENGVITLS AQPWGEKALIQVTDGCGIPDFALPEIFDFRYSLPRENGKSSGGLAFVSEARLLN GEVTLGNREGEGLASLTLRHPT"

2812..4173

/gene="cred"

/note="synonym: STM4590"

2812..2817

/gene="cred"

/note="putative RBS for cred; RegulonDB:STMSIH004450"

2824..4173

/gene="cred"

/note="similar to E. coli tolerance to colicin E2 (AAC77353.1); Blastp hit to AAC77353.1 (450 aa), 73%

identity in aa 1 - 450"

/codon_start=1

/transl_table=11

/product="tolerance to colicin E2 protein"

/protein_id="AAL23405.1"

/db_xref="GI:16423157"

/translation="MLKSPILMKMTITLGGAMTLLILPLMMVHTTVERADYSHVEA IKQSTSPQKVVGPVIAIPVTELYTLVEEKEVQKRSYLFWLPESLVEGNQVNEA RIGTIYQGVMTDPAIKAEFDVARLHELRNITLTKPFIIVGVGDARGLSAVKAPO VNGEITLVPGTGLPESRGGIHLPLPDSQMATRNLTLDKSLNLSGGSFSLVPGRSS EKTLSNMPHPNPNVDFPLGKREISGSGQAGQWOTSRFATNIGEPADAOXKDMDLPS AFSVAVSTPADDYQUTDRATKRAIILITITTPAFVFFETLTCRLHPMYTLYVGLSLV MRYLLALASERHIGFTPAWIASLVGALMNSVYLQAVLQGMNSVLPTVALALADGVW MGLRSDESSLIGTGVLLALAGVWFLTRHLDWYLSQCKRKSLLPPVXDDELRLMK"

```
/gene="sthd"
/notes="similar to E. coli fibrial morphology
(AAC7275.1); Blastp hit to AAC7275.1 (167 aa), 31%
identity in aa 18 - 166"
/codon_start=1
/translation=1
/product="putative fibrial subunit"
/protein_id="AAL23407.1"
/db_xref="GI:16423159"
/translation="NMNRFPALALITPTGLMSPSIAAYTADAVGINVGLIY
ADPCIVPEPTLSIGQYARADJAVAGNATPTPTKMCNCAATTTKAIIFNPGYP
RMYTPHGXTPAGDFISVITITAMQ"
complement(5924..5929)
/gene="sthd"
/notes="putative RBS for sthd; RegulonDB:STM51H004452"
```

RBS

```
Query Match 77.0%; Score 714; DB 1; Length 13852;
Best Local Similarity 92.1%; Pred. No. 6,1e-201;
Matches 753; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 GTGATGTTACGCCGATCATGTTAATTGAGCATCATCAGCAGGTGAGGACTTTTGT 60
DB 12606 GTGATGTTACGCCGATCATGTTAATTGAGCATCATCAGCAGGTGAGGACTTTTGT 12547
QY 61 ACTTCCTGTTGATTTAGTTGGCAATTAGTAGCAACATGCAACCCCGACATTTCT 120
DB 12546 ACTTCCTGTTGATTTAGTTGGCAATTAGTAGCAACATGCAACCCCGACATTTCT 12487
QY 121 TATCGTTGAACGAGTTGTTAACAACGACACGTTGAAAAGTATTTTTCGACGGGAAG 180
DB 12486 TATCGTTGAACGAGTTGTTAACAACGACACGTTGAAAAGTATTTTTCGACGGGAAG 12427
QY 181 CTATGATGTTTTCGACGACAGATGCGCGGAAATGATCATGATCTCTTGAATATGA 240
DB 12426 CTATGATGTTTTCGACGACAGATGCGCGGAAATGATCATGATCTCTTGAATATGA 12367
QY 241 CATCAACCTGTTGATGATGATATCATCTGCGGGTGAAGACGTTCTGTTACCGG 300
DB 12366 CATCAACCTGTTGATGATGATATCATCTGCGGAAACGTTCTGTTACCGG 12307
QY 301 TGAATGCGGAGAGCGGAAATGTTGTTGATGTTCTGATCGCGGTGAACAGAGT 360
DB 12306 TGAATGCGGAGAGCGGAAATGTTGTTGATGTTCTGATCGCGGTGAACAGAGT 12247
QY 361 CGATTAATTTCTGCGCTCGAAATCGGATGATGATCATCATCAACAAACGTTCAACC 420
DB 12246 CGATTAATTTCTGCGCTCGAAATCGGATGATGATCATCATCAACAAACGTTCAACC 12187
QY 421 GCGTGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 12186 GCGTGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12127
QY 481 CAGCGAAGAACGTCGTAGGCTTGAAGCTACAAAGTTCAATGTTGGAACCTGACATGAA 540
DB 12126 CAGCGAAGAACGTCGTAGGCTTGAAGCTACAAAGTTCAATGTTGGAACCTGACATGAA 12067
QY 541 CAGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 12066 CAGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12007
QY 601 CCGCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCGTGTGAACCTGTGAA 660
DB 12006 CCGCATGCTTCACTTCTGTGAAACCCAGGCAAAATTCAGTCCGTGTGAACCTGTGAA 11947
QY 661 GAAATGACCGGCGGTGAGCTGAAACCGCAGACGATGATGATGATGATGATGATGATGAT 720
DB 11946 GAAATGACCGGCGGTGAGCTGAAACCGCAGACGATGATGATGATGATGATGATGATGAT 11897
QY 721 TATGTTAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 11896 TATGTTAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11827
```

```
QY 781 TGAAGTTATCGCTTCGCGGTATCTGGAAGATTAAT 818
DB 11826 CGAAGTTATCGCTTCGCGGTATCTGGAAGATTAAT 11789
```

```
RESULT 13
AR385083
LOCUS AR385083 726 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1812 from patent US 6610836.
ACCESSION AR385083
VERSION AR385083.1 GI:40094817
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 726)
AUTHORS Breton G.L. and Osborne M.
TITLE Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
for diagnostics and therapeutics
JOURNAL Patent: US 6610836-A 1812 26-Aug-2003;
FEATURES
source location/Qualifiers
1..726
/organism="unknown"
/mol_type="genomic DNA"
```

ORIGIN

```
Query Match 62.3%; Score 577.2; DB 6; Length 726;
Best Local Similarity 87.2%; Pred. No. 2,1e-160;
Matches 633; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

```
QY 92 GTAGCAACATGCAACCGCCGACATTTCTTATGTTGAAGACGATGTTGAACGCAAC 151
DB 1 GTAGCAACATGCAACCGCCGACATTTCTTATGTTGAAGACGATGTTGAACGCAAC 60
QY 152 ACGTTGAAAGTATTTTTCGAAAGCGAGGCTATGATGTTTGAAGCGACATGCGCG 211
DB 61 ACGTTGAAAGTATTTTTCGAAAGCGAGGCTATGATGTTTGAAGCGACGATGCGCG 120
QY 212 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 271
DB 121 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 272 CCGGGTGAAGACCGTTCTGTTAGCGGTGAACGCGCGAGCGCAATGTTGCGTTG 331
DB 181 CCGGGTGAAGACCGTTCTGTTAGCGGTGAACGCGCGAGCGCTGAAGTGGCGCTG 240
QY 332 ATGTTCTGATCGCGCGGTGACAAAGATGATGATGATGATGATGATGATGATGATGAT 391
DB 241 ATGTTCTGATCGCGCGGTGACAAAGATGATGATGATGATGATGATGATGATGATGAT 300
QY 392 GATGACTATCATCAACCAACCGTTCAACCGCGGTGAACGATGATGATGATGATGATGAT 451
DB 301 GATGACTATCATCAACCAACCGTTCAACCGCGGTGAACGATGATGATGATGATGATGAT 360
QY 452 CTGTCCCGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511
DB 361 CTGTCCCGTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 512 AAGTTCAATGTTGGAACCTGATCATCAACCGCTTCTGTTATCGGCTGTGATGAG 571
DB 421 AAGTTCAACCGCTTCTGATCATCAACCGCTTCTGTTATCGGCTGTGATGAG 480
QY 572 CAGTGAAGCTGCGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 631
DB 481 CAGTGAAGCTGCGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 632 AAAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 691
DB 541 AAAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 692 GACCGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 751
DB 601 GACCGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
```


QY 752 ACCCGGAAATCATCGCCACCATTCACGGTGAAGGTTATCGCTTCGCGGTGATCTGGAA 811
DB 661 ACCCGGAAATCATCGCTACCATTCACGGTGAAGGTTATCGCTTCGCGGTGATCTGGAA 720
QY 812 GATTAA 817
DB 721 GAATTA 726

RESULT 14
AF267619 564 bp DNA linear BCT 03-AUG-2000
LOCUS Escherichia coli strain E2348/69 aerobic regulator (arca) gene,
DEFINITION Partial cds.
ACCESSION AF267619
VERSION AF267619.1 GI:9664513
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 564)
AUTHORS Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and
TITLE Parallel evolution of virulence in pathogenic Escherichia coli
JOURNAL Nature 406 (6791), 64-67 (2000)
MEDLINE 20351039
PUBMED 10894541
REFERENCE 2 (bases 1 to 564)
AUTHORS Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and
TITLE Direct Submision
JOURNAL Submitted (12-MAY-2000) Biology, Pennsylvania State University, 208
FEATURES
source
1. 564
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="E2348/69"
/serotype="O127:H6"
/specimen_voucher="Taunton 69"
/db_xref="taxon:562"
/note="EPEC 1"
/gene="arca"
/gene_start=1
/codon_start=1
/transl_table=11
/product="aerobic regulator"
/protein_id="AAF97159.1"
/db_xref="GI:9664513"
/translation="ATDGAEMHQILSEYDINIVIMDIN.PGKNGLLIAREQANVA
LMEULTRDNEVDKILGLEIGADVDYIKPPEPLTIRARLISRINVLGVSEERSV
ESYKFGWELDINSRLIGDGEQYKLPREPFAMHFCENPESKIOSRAELIKMTGR
ELKPHRDVTVTRIRIKHPESPDPPE"

ORIGIN
Query Match 60.7%; Score 562.4; DB 1; Length 564;
Best Local Similarity 99.8%; Pred. No. 5,2e-156;
Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 197 GCGACGATGGCGCGGAAATGCATGCATCTCTGAAATATGACATCAACCTGGGATC 256
DB 1 GCGACGATGGCGCGGAAATGCATGCATCTCTGAAATATGACATCAACCTGGGATC 60
QY 257 ATGATATCATCTGCGCGGTAAAGAACGATCTTCTGTTAGCGGTGAATGCGCGAGCAG 316
DB 61 ATGATATCATCTGCGCGGTAAAGAACGATCTTCTGTTAGCGGTGAATGCGCGAGCAG 120
QY 317 GCGAATGTTGCGTTGATGTTCTGACTGCGCGGTGACAGCAAGTCGATTAATTTCTGGC 376

DB 121 GCGAATGTTGCGTTGATGTTCTGACTGCGCGGTGACAGCAAGTCGATTAATTTCTGGC 180
QY 377 CTCGAATCGGTGCAAGTACTACATCACCAACCGTTCAACCGCGTGAATGACGAT 436
DB 181 CTCGAATCGGTGCAAGTACTACATCACCAACCGTTCAACCGCGTGAATGACGAT 240
QY 437 CGTGACGCAACTACTGTCCTGTTACCATGATCTGGGTACTGTACGCGAAGACGTGT 496
DB 241 CGTGACGCAACTGCTGTCCTGTTACCATGATCTGGGTACTGTACGCGAAGACGTGT 300
QY 497 AGCGTGAAGCTACCAAGTTCAATGTTGGGAACTGCACTCAACACCGCTTGTGATC 556
DB 301 AGCGTGAAGCTACCAAGTTCAATGTTGGGAACTGCACTCAACACCGCTTGTGATC 360
QY 557 GCGCCTGATGCGCAGCAGTACAGCTGCGCGCAGCAGGATTCGCGCCATGTTCACTTC 616
DB 361 GCGCCTGATGCGCAGCAGTACAGCTGCGCGCAGCAGGATTCGCGCCATGTTCACTTC 420
QY 617 TGTGAAAACCCAGCAAAATTCAGTCCCTGCTGAACTGCTGAAGAAAATGACCGCGCT 676
DB 421 TGTGAAAACCCAGCAAAATTCAGTCCCTGCTGAACTGCTGAAGAAAATGACCGCGCT 480
QY 677 GAGCTGAACCCGACGACCGCTACTGTAGACGTACGATCCGCGTATTCGTAACATTTTC 736
DB 481 GAGCTGAACCCGACGACCGCTACTGTAGACGTACGATCCGCGTATTCGTAACATTTTC 540
QY 737 GAATCTACGCGCGATACGCGCGAA 760
DB 541 GAATCTACGCGCGATACGCGCGAA 564

RESULT 15
AF267620 564 bp DNA linear BCT 03-AUG-2000
LOCUS Escherichia coli strain DECia aerobic regulator (arca) gene,
DEFINITION Partial cds.
ACCESSION AF267620
VERSION AF267620.1 GI:9664515
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 564)
AUTHORS Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and
TITLE Parallel evolution of virulence in pathogenic Escherichia coli
JOURNAL Nature 406 (6791), 64-67 (2000)
MEDLINE 20351039
PUBMED 10894541
REFERENCE 2 (bases 1 to 564)
AUTHORS Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and
TITLE Direct Submision
JOURNAL Submitted (12-MAY-2000) Biology, Pennsylvania State University, 208
FEATURES
source
1. 564
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="DECia"
/serotype="O5:H6"
/specimen_voucher="CDC 572-56"
/db_xref="taxon:562"
/note="EPEC 1"
/gene="arca"
/gene_start=1
/codon_start=1
/transl_table=11
/product="aerobic regulator"
/protein_id="AAF97160.1"

/db_xref="GI:9664516"
/translation="ATDGAEMHOILSEYDINLVIMDINLPKNGILLARELEQANVA
LMLTGRDNEVDKILGLEIGADDYITKPEVNPRLTIRARNLISRTNLCITVSEERSV
ESYKFNEMELDINSRLIGPDSEYKILPRSEPRAMLHPCENPKLIOSRAELIKMTGR
ELKPHDRTVIVTIRIRIKHESTPDPPE"

ORIGIN

Query Match 60.7%; Score 562.4; DB 1; Length 564;
Best Local Similarity 99.8%; Pred. No. 5.2e-156;
Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 197 GCGACAGATGGCGGGAATGATCATCTCTGAAATATGACATCAACTGGTGATC 256
DB 1 GCGACAGATGGCGGGAATGATCATCTCTGAAATATGACATCAACTGGTGATC 60
QY 257 ATGATATCAATTCGCGGGTAAGAACGGTCTTCTGTTAGCGGTGAACCTGGCGAGCAG 316
DB 61 ATGATATCAATTCGCGGGTAAGAACGGTCTTCTGTTAGCGGTGAACCTGGCGAGCAG 120
QY 317 GCGAATGTTGCGTGTGATGTTCTGGACTGGCCGTGACAAAGATGATTAATTCGCGC 376
DB 121 GCGAATGTTGCGTGTGATGTTCTGGACTGGCCGTGACAAAGATGATTAATTCGCGC 180
QY 377 CTGGAATGGTGCAGATGACTACACAAACCGTCAACCGCGTGAACGATC 436
DB 181 CTGGAATGGTGCAGATGACTACACAAACCGTCAACCGCGTGAACGATC 240
QY 437 CGTGACGCAACCTGCTGCTCCGCTACCAATCTGGGTACTGTCAAGCAAGAACTGCGT 496
DB 241 CGTGACGCAACCTGCTGCTCCGCTACCAATCTGGGTACTGTCAAGCAAGAACTGCGT 300
QY 497 AGCGTTGAAGCTACAACTTCATGTGTGGAACTGGACATCAACAGCCGTTGATC 556
DB 301 AGCGTTGAAGCTACAACTTCATGTGTGGAACTGGACATCAACAGCCGTTGATC 360
QY 557 GAGCCCTGATGGGAGAGCTGACAGCTCCGCGCAGCGAGTTCCGCGCATCTTCACTTC 616
DB 361 GAGCCCTGATGGGAGAGCTGACAGCTCCGCGCAGCGAGTTCCGCGCATCTTCACTTC 420
QY 617 TGTGAAAAACCGCAAAATTCAGTCCGCTGCTGAAGAAATGACCGGCGGT 676
DB 421 TGTGAAAAACCGCAAAATTCAGTCCGCTGCTGAAGAAATGACCGGCGGT 480
QY 677 GAGCTGAACCGCAGCAGCCTGTAAGTACGTGACGATCCGCGTATTCGTAACATTTTC 736
DB 481 GAGCTGAACCGCAGCAGCCTGTAAGTACGTGACGATCCGCGTATTCGTAACATTTTC 540
QY 737 GAATCTAGCCCGGATAGCCCGGAA 760
DB 541 GAATCTAGCCCGGATAGCCCGGAA 564

Search completed: January 22, 2005, 13:53:24
Job time : 4273 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 10:44:08 ; Search time 551 Seconds
(without alignments)
831.603 Million cell updates/sec

Title: US-10-613-990A-31

Sequence: 1 gcatcgtcaccgcgcacatcatg.....gtactcttactgcttc 927

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_23sep04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	927	100.0	927	12	ADJ38402
2	925.8	99.9	1732	6	ABL90421
3	577.2	62.3	726	11	ACH96017
4	510.6	55.1	110000	10	ACF67367_28
5	510.6	55.1	110000	10	ACF65386_0
6	508.6	54.9	759	12	ADJ38380
7	493	53.2	717	10	ACF69844
8	450.8	48.6	726	10	ADP01748
9	447.4	48.3	1048	5	AAS29239
10	319.4	34.5	110000	2	AAT42063_09
11	316.2	34.1	711	4	AAS3409
12	316.2	34.1	711	8	ACA34287
13	232.8	25.1	1110	5	AAS82422
14	137.4	14.8	729	6	ABQ90117
15	126	13.6	837	11	ABD17813
16	126	13.6	2430	11	ABD17567
17	126	13.6	3675	11	ABD17705
18	120.6	13.0	990	11	ABD12913
19	120.6	13.0	2166	11	ABD13118
20	113.8	12.3	2004	11	ABD13261
21	107.4	11.6	1018	3	AAA38554

C	22	106.8	11.5	11275	6	AAH43717	Aah43717 E. coli g
	23	103.8	11.2	714	8	ACA47574	Aca47574 Prokaryot
	24	101.4	10.9	705	8	ACA48118	Aca48118 Prokaryot
	25	100	10.8	100	8	ACD69947	Acdd69947 E. coli K
	26	100	10.8	100	8	ACD69946	Acdd69946 E. coli K
	27	100	10.8	100	8	ACD69948	Acdd69948 E. coli K
	28	99	10.7	732	8	ACA26627	Aca26627 Prokaryot
	29	98.6	10.6	110000	6	ABA03041_03	Continuation (4 of
	30	98	10.6	714	8	ACA36700	Aca36700 Prokaryot
	31	97.2	10.5	762	11	ABD07703	Abd07703 Pseudomon
C	32	96.8	10.4	349980	6	ABQ81844	Abq81844 Bifidobac
	33	96.2	10.4	705	6	ABN90763	Abn90763 Staphyloc
	34	96.2	10.4	721	8	ACA46767	Aca46767 Prokaryot
	35	95	10.2	110000	6	ABQ69245_03	Continuation (3 of
	36	95	10.2	110000	6	ABQ67197_02	Continuation (3 of
	37	94.6	10.2	702	4	AAH52543	Aah52543 S. epider
	38	94.6	10.2	3389	4	AAH54550	Aah54550 S. epider
	39	93	10.0	699	4	AAS52039	Aas52039 Staphyloc
	40	93	10.0	702	3	AAH55530	Aah55530 S. aureus
	41	93	10.0	708	4	AAS4939	Aas4939 Staphyloc
	42	93	10.0	744	8	ACA25690	Aca25690 Prokaryot
	43	92.6	10.0	792	11	ABD11939	Abd11939 Pseudomon
	44	92.6	10.0	804	11	ABD11719	Abd11719 Pseudomon
	45	91.4	9.9	699	4	AAS00812	Aas00812 S. aureus

ALIGNMENTS

RESULT 1	ADJ38402	standard; DNA; 927 BP.
ID	ADJ38402	
XX	ADJ38402;	
AC		
XX		
DT	06-MAY-2004	(first entry)
XX		
DE	Escherichia coli arca (aerobic respiration control) gene SeqIDJ1.	
XX		
KW	gamma-proteobacterium; Arca; aerobic respiration control;	
KW	Arca protein function; chromosomal arca gene; fermentation; L-amino acid;	
KW	L-lysine; L-glutamic acid; gene; de.	
XX		
OS	Escherichia coli.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	101..817
FT		/tag= a
FT		/product= "Escherichia coli arca protein"
XX		
PN	EP1382666-A1.	
XX		
PD	21-JAN-2004.	
XX		
PF	11-JUL-2003; 2003EP-00015911.	
XX		
PR	12-JUL-2002; 2002JP-00203764.	
XX		
PA	(AJIN) AJINOMOTO CO INC.	
XX		
PI	Ishikawa Y, Imazumi A, Matsui K, Kojima H;	
DR	WPI; 2004-192975/19.	
DR	P-PSDB; ADJ38403.	
XX		
PT	New bacterium useful for producing a target substance (especially L-	
PT	lysine, L-glutamic acid or another L-amino acid) is modified so that the	
XX	Arca protein does not normally function.	
XX		
PS	Claim 6; SEQ ID NO 31, 38pp; English.	
XX		
CC	This invention relates to a novel gamma-proteobacterium able to produce a	
CC	target substance such as an amino acid which is modified so that an Arca	

Db	1313	GTCAATGTTAGCCCGCATCTAATGTTAATTGTCAGATGATCATCAGACAGGTCAAGGACCTTTTGT	125
Oy	61	ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAATGACAGACCCCGCACATCT	120
Db	1253	ACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAAATGACAGACCCCGCACATCT	1194
Oy	121	TATGTTGAACACGAGTGGTGAACACGGAACACGTTGAAGTAATTTTCGAAGCGGAAG	180
Db	1193	TATGTTGAACACGAGTGGTGAACACGGAACACGTTGAAGTAATTTTCGAAGCGGAAG	1134
Oy	181	CTATGATGTTTTCGAAAGCAGACAGATGGCGCGAAATGATCATGATCCTCTCGAATATGA	240
Db	1133	CTATGATGTTTTCGAAAGCAGACAGATGGCGCGAAATGATCATGATCCTCTCGAATATGA	1074
Oy	241	CATCAACCTGCTGATCATGATATCAATCTGCCGGGTAAAGACGGTCTTTCTGTAGCCG	300
Db	1073	CATCAACCTGCTGATCATGATATCAATCTGCCGGGTAAAGACGGTCTTTCTGTAGCCG	1014
Oy	301	TGAACCTGGCGCAGACGGGGAATGTTTGCTTGAATGTTCTTGACTGGCCGTGCAACGAAGT	360
Db	1013	TGAATGTCGACAGCAGGCGAAATGTTTGCTTGAATGTTCTTGACTGGCCGTGCAACGAAGT	954
Oy	361	CGATPAAATTTCTCGGCTTCGAATACTGGGTGACAGATACATCAACAAACGGTTCAACC	420
Db	953	CGATPAAATTTCTCGGCTTCGAATACTGGGTGACAGATACATCAACAAACGGTTCAACC	894
Oy	421	GCGTGAACCTGACGATTCGTGACCGCAACTACTGTCCCGTACCATGAATCTGGTACTGT	480
Db	893	GCGTGAACCTGACGATTCGTGACCGCAACTACTGTCCCGTACCATGAATCTGGTACTGT	834
Oy	481	CAGCGAAGAACGTCGTAGCGTTGAAAGCTACAGTTCAATGTTGGGAACTGGACATCA	540
Db	833	CAGCGAAGAACGTCGTAGCGTTGAAAGCTACAGTTCAATGTTGGGAACTGGACATCA	774
Oy	541	CAGCGGTCGTTGATCGGCGCCGTGAATGGGAGAGTACAAAGCTCGCGCAGCGAGTTCCG	600
Db	773	CAGCGGTCGTTGATCGGCGCCGTGAATGGGAGAGTACAAAGCTCGCGCAGCGAGTTCCG	714
Oy	601	CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA	660
Db	713	CGCCATGCTTCACTTCTGTGAAAAACCCAGGCAAAATTCAGTCCCGTGTGAACTGCTGAA	654
Oy	661	GAAATATGACCGGCGGTGAGCTGAAACCGCACGACCTGACTGTAAGCTGACGATCCGCG	720
Db	653	GAAATATGACCGGCGGTGAGCTGAAACCGCACGACCTGACTGTAAGCTGACGATCCGCG	594
Oy	721	TATTCGTAAATTTTGAATCTAGCCGGATACAGCCGGAATTCATCCGACCACTTACAGG	780
Db	593	TATTCGTAAATTTTGAATCTAGCCGGATACAGCCGGAATTCATCCGACCACTTACAGG	534
Oy	781	TGAAGGTTATCGCTTCTGCGGTGATCTGGAAGATTAATCGGCTTTACACCGTCAAAAA	840
Db	533	TGAAGGTTATCGCTTCTGCGGTGATCTGGAAGATTAATCGGCTTTACACCGTCAAAAA	474
Oy	841	AACGGCGCTTTTAAAGCGCGCTTTTAAATTTTCAACTTATTTCCAGATAGCTAACTATC	900
Db	473	AACGGCGCTTTTAAAGCGCGCTTTTAAATTTTCAACTTATTTCCAGATAGCTAACTATC	414
Oy	901	GTCGCTTGAATCTTTACTGCGCTTT	927
Db	413	GTCGCTTGAATCTTTACTGCGCTTT	387

	RESULT 3
ACH96017	
ID	ACH96017 standard; DNA; 726 BP.
XX	
XX	ACH96017;
AC	
XX	
Dt	29-JUL-2004 (first entry)
XX	
DE	Klebsiella pneumoniae polynucleotide seqid 1812

Query Match	Best Local Similarity	Matches	Score	DB	Length
62.3%;	87.2%;	0;	577.2;	11;	726;
Conservative	0;	Mismatches	93;	Indels	0;
Gaps	0;				
Sequence 726 BP; 189 A; 198 C; 189 G; 150 T; 0 U; 0 Other;					
Recombinant expression vector; transcription regulatory element;					
Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.					
Klebsiella pneumoniae.					
US6610836-B1.					
26-AUG-2003.					
27-JAN-2000; 2000US-00489039.					
29-JAN-1999; 99US-0117747P.					
(GENO-) GENOME THERAPEUTICS CORP.					
Breton GL, Osborne M;					
WPI: 2003-895346/82.					
P-PSDB; ABO62466.					
New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for preparing a vaccine composition against Klebsiella pneumoniae.					
Disclosure; SEQ ID NO 1812; 932pp; English.					
The invention describes a new isolated nucleic acid encoding a Klebsiella pneumoniae polypeptide. Also described are: a recombinant expression vector comprising the nucleic acid, operably linked to a transcription regulatory element; and a cell comprising the recombinant expression vector. The nucleic acid is useful for preparing a vaccine composition against Klebsiella pneumoniae. This sequence encodes a Klebsiella pneumoniae polypeptide of the invention					
92 GTAGCAAAACATGACGAGACCCCGACATTTCTTATCGTTGAAGAAGAGTTGGTAACACGCAAC	151				
1 GTAGCAAAACATGACGAGACCCCGACATTTCTTATCGTTGAAGAAGAGTTGGTAACACGCAAC	60				
152 ACGTTGAAAAGTATTTTCGAAAGCGGAAGGCTATGATGTTTTCGAAGCAGACAGATGGCGCG	211				
61 ACGTTAAAAGTATTTTCGAAAGCGGAAGGCTATGATGTTTTCGAAGCAGACAGATGGCGCG	120				
212 GAAATGCATCAGATCCTCTCTGGAATATGACATCAACTGGTGATCATGATATCAATCTG	271				
121 GAAATGCATCAGATCCTCTCTGGAATATGACATCAACTGGTGATCATGATATCAATCTG	180				
272 CCGGGTAAAGAACGGTCTTCTGGTTGGCGGCTGTAACCTGGCGGACGGAATGTTGCGTTG	331				
181 CCGGGTAAAGAACGGTCTTCTGGTTGGCGGCTGTAACCTGGCGGACGGAATGTTGCGTTG	240				
332 ATGTTTCCGATCGGCGCGTGAACAGAGTGCATATAATTTCTGGCGCTCGAATCGGTGCA	391				
241 ATGTTTCCGATCGGCGCGTGAACAGAGTGCATATAATTTCTGGCGCTCGAATCGGTGCA	300				
392 GATGACATACATACCAAAACCGTTCAACCCGCGTGAATGACGATTCGTGCAACGCAACTTA	451				
301 GATGACATACATACCAAAACCGTTCAACCCGCGTGAATGACGATTCGTGCAACGCAACTTA	360				
452 CTGTCGCCGTAACGAATCTGGGGTATCTGACGGAAGAACGTCGATGAGGTTGGAAGGCTAC	511				
361 CTGTCGCCGTAACGAATCTGGGGTATCTGACGGAAGAACGTCGATGAGGTTGGAAGGCTAC	420				
512 AAGTTCAATGAGTTGGGAACTGGAATCAACAGCCGTTCTGTTGATCGAGCCCTGATGGCGAG	571				
421 AAGTTCAATGAGTTGGGAACTGGAATCAACAGCCGTTCTGTTGATCGAGCCCTGATGGCGAG	480				
572 CAGTACAAAGCTGCGCGCAGGAGGTTCCGCGCATGCTTCACTTCTGTGAAAACCCAGGC	631				

MP	1	110000
MP	ACF67367_00	210000
MP	ACF67367_01	100001
MP	ACF67367_02	200001
MP	ACF67367_03	300001
MP	ACF67367_04	400001
MP	ACF67367_05	500001
MP	ACF67367_06	600001
MP	ACF67367_07	700001
MP	ACF67367_08	800001
MP	ACF67367_09	900001
MP	ACF67367_10	1000001
MP	ACF67367_11	1100001
MP	ACF67367_12	1200001
MP	ACF67367_13	1300001
MP	ACF67367_14	1400001
MP	ACF67367_15	1500001
MP	ACF67367_16	1600001
MP	ACF67367_17	1700001
MP	ACF67367_18	1800001
MP	ACF67367_19	1900001
MP	ACF67367_20	2000001
MP	ACF67367_21	2100001
MP	ACF67367_22	2200001
MP	ACF67367_23	2300001
MP	ACF67367_24	2400001
MP	ACF67367_25	2500001
MP	ACF67367_26	2600001
MP	ACF67367_27	2700001
MP	ACF67367_28	2800001
MP	ACF67367_29	2900001
MP	ACF67367_30	3000001
MP	ACF67367_31	3100001
MP	ACF67367_32	3200001
MP	ACF67367_33	3300001
MP	ACF67367_34	3400001
MP	ACF67367_35	3500001
MP	ACF67367_36	3600001
MP	ACF67367_37	3700001
MP	ACF67367_38	3800001
MP	ACF67367_39	3900001
MP	ACF67367_40	4000001
MP	ACF67367_41	4100001
MP	ACF67367_42	4200001
MP	ACF67367_43	4300001
MP	ACF67367_44	4400001
MP	ACF67367_45	4500001
MP	ACF67367_46	4600001
MP	ACF67367_47	4700001
MP	ACF67367_48	4800001
MP	ACF67367_49	4900001

44 AGCGCAGGGAATTGTTGACTTCCCTGTTTGCAATTTAGTTAGGCAATTTAGTAGAGCAAAATG 103
Db 67943 AGATTAAAGTAGTTGATTGAATCAATTTCGATTTATTTTTTAAAGCATTCTAGGTGCAACCATG 68001
Qy 104 CAGACCCCGACATTCCTTATTCGTGTAAGACGAGTTGGTATGACCAACCACTTGAAAGT 163
Db 68003 CAAACCCCGACATTTTGAATTGTTGTAAGACAAATTGTCACCTCCCAATPACCTTAAAGC 68067
Qy 164 ATTTTGAACGGAAAGGCTATGATGTTTTCGAAGCAGATGGCGGGAATGATCAAG 223
Db 68063 ATTTTGAACGTAAGGGTACATGATTATGAAAGCAGTATGTTTGAATAATGCAACAT 68122
Qy 224 ATCCTCTGGAATATATACATCAACTGGTATCAATGATATCAATCTGCCGGTAAGAC 283
Db 68123 ATTCTGTCAAAACAGACATTTTAACTCGGTGATTTGAGATTTTAACTTCCAGGCAAAAT 68182
Qy 284 GGTCTTCTGTAGCGCGTGAACCTGCGCAGACAGCGCAATGTTGCGTTGATGTTCCGACT 343
Db 68183 GGTCTTGTACTTGGCCCGCAACTGGAGAGCAACCAATGTCTTGAATGTTCGACT 68242
Qy 344 GCGCGGAACAAGATGATTAATAATTCGCGCTCGAAATCGTGCAGATGATACATC 403
Db 68243 GCGCGGATTAAGAAATGAGACAAATCTTGGCTTGGAAATCGGTGCAATGATTACATC 68302
Qy 404 ACCAAACCGTTTCAACCCGCGTGAACCTGACATTCGTGACGCAACCTATGTCGCCGTACC 463
Db 68303 ACCAAACCATTTAATCCCGCGTAATTGACTATTCGCGCCCGCAACTGCTTCTGTGACC 68362
Qy 464 ATGAATCTGGGTAAGTGTACGAGCAAGAACGTCGTAGCGTTGAAAGCTACAAAGTCAATGGT 523
Db 68363 ATGAATCTGAAGTATGTCAAGAGAAACGTCGCAAGTTGAGAGTTATTAATTCAACGGT 68422
Qy 524 TGGGAATCGGCAATCAACACCCGTTGCTGATTCGACCTGATGGCGAGCAATGACAGCTG 583
Db 68422 TGGGAATTAGTATCAATACCGCTTCTGATTAAGCCGCGCTGTGAGCCGATTAATTG 68482
Qy 584 CCGCGCAGCAGATTCCGCGCATAGTCTCACTCTGTGAAAACCCAGGGGAAAATTCAGTCC 643
Db 68483 CCAAGCAGCAGATTCCGCGCATAGTCTCACTCTGTGAAAATCCAGGGGAAAATCCAAACA 68542
Qy 644 GGTCTGAACCTGCTGAAGAAATGACCGGCGGTGAGCTGAAGACGACGACCTGACTGA 703
Db 68543 CGCCCAATTTTACTGAAAAAGATGACAGGTGCTGAATTTGAACCTCAAGACCGTACTGTT 68602
Qy 704 GACGTGACGATTCGCGGTATTCGTGAATCAATTTCCGAATCTTACGCCGATACGCCGGAAATC 763
Db 68603 GATTTTAAACGATTTGTTGTAATCCGTAAACATTTTGAATGAGCGCAGATACGATGAAAT 68662
Qy 764 ATGCGCAACCAATTACCGTGAAGGTTATCGCTTTCGCGGTATCTGGAAGATTAATGGCT 823
Db 68663 ATGCGCAACCAATTACCGTGAAGGCTATCGCTTTCGTGTGATTTGGGAAGATTAACCTGAA 68722
Qy 824 TTACCAACGTCAGAAA 840
Db 68723 CTATCGGTCTTAACA 68729

RESULT 5
ACF65386_0/c
WP Sequence split into 7 fragments LOCUS ACF65386 Accession ACF65386

WP Fragment Name Begin End
 WP ACF65386_0 1 110000
 WP ACF65386_1 100001 210000
 WP ACF65386_2 200001 310000
 WP ACF65386_3 300001 410000
 WP ACF65386_4 400001 510000
 WP ACF65386_5 500001 610000
 WP ACF65386_6 600001 700779
 ID ACF65386 standard; DNA; 700779 BP.
 AC ACF65386;
 XX
 XX 20-NOV-2003 (first entry)
 DE
 XX Photorhabdus luminescens nucleotide sequence #39.
 XX
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KM detection; food; gene expression; plant; animal; microorganism; toxin;
 KM antibiotic; bioplastic; virulence factor; disease model; plague;
 XX whooping cough; gene; ds.
 XX
 OS Photorhabdus luminescens.
 XX
 PN MO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002MO-IB003040.
 XX
 PR 07-FEB-2001; 2001PR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;
 PI Buchrieser C;
 DR MPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides.
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 1; SEQ ID NO 39; 1205BP; French.
 XX
 XX The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi) that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens genes
 CC
 SQ Sequence 700779 BP; 190440 A; 140977 C; 164444 G; 204917 T; 0 U; 1 Other;
 Query Match 55.1%; Score 510.6; DB 10; Length 110000;
 Best Local Similarity 77.5%; Pred. No. 1e-157;
 Matches 618; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
 QY 44 AGGTACGGAACCTTTGACTTCTGTTGCAATTAGTTCGCAATTAGAGCAAAACATG 103

DB 30684 AGATAAGAGTAGTTGTAATCAATTCGATTATTTTAACGATTTAGAGCAACCATG 30625
 QY 104 CAGACCCCGCAGCATTTCTTATCGTTGAAGAGAGAGTTGTAACAAGCAACAGTTGAAAGT 163
 DB 30624 CAAACCCCGCAGCATTTGATTTGTTGAAGAGAGAAATGTCAGTCCGATATCCCTAAAAAGC 30565
 QY 164 ATTTTGAAGCGGAGGCTATGATGTTTTCGAAAGCAAGAGTGGCCGGAATGATCAG 223
 DB 30564 ATTTTGAAGCTGAAGGGTATCAATGATTATGAAGCCATGATGTTGAGAAATGACACAT 30505
 QY 224 ATCTCTCTGAATATATCAATCAACCTGATGATCATGATATCAATGATGCGGTAAGAC 283
 DB 30504 ATTTCTGCAAAACAGCAATTAACCTGATGATTTATGATTAATTAACCTTCAGGCAAAAT 30445
 QY 284 GGCTCTCTGTTAGCGGCTGAACCTGCGAGAGAGGCAATGTTGCTGTTGATCTGACT 343
 DB 30444 GGCTTCTTACTTCCCGCACTGCGAGAGAGCAAAATGTTGCTTGAATGTTCTGACT 30385
 QY 344 GGCGGTGACAAAGCAAGTCAATAAATTCGCTCGAAATCGGTCAAGTCAATCATC 403
 DB 30384 GGCGGTGATTAAGCAAGTCAATAAATTCGCTCGAAATCGGTCAAGTCAATCATC 30325
 QY 404 ACCAAACCGTTCAACCCCGGTGAACGATTCGTCACGCAACCTACTGTCCTGATCC 463
 DB 30324 ACCAAACCAATTAATCCCGGTGAATGATTAATGATTCGCTGCTTCTGATCC 30265
 QY 464 ATGATCTGGGTACTGTCAGAGCAAGAGCGTGAAGGCTGGAAGTCAATGATCAATGAT 523
 DB 30264 ATGATCTGATGATATTCAGAGCAAGAGCGTGAAGGCTGGAAGTCAATGATCAATGAT 30205
 QY 524 TGGGAATGGAATCAACAGCGGTTGATCGGCTGATGAGCGAGAGTCAACAGT 583
 DB 30204 TGGGAATGGAATCAATGATCGGCTGATGAGCGGCTGATGAGCGTGAATGAT 30145
 QY 584 CCGCGAGAGGATTCGCGCCGATGCTTCACTTGTGAAAAACCGAGCAAAATTCAGTCC 643
 DB 30144 CCGCGAGAGGATTCGCGCCGATGCTTCACTTGTGAAAAATTCAGCAAAATTCAGTCC 30085
 QY 644 CGTGTGAATCTGCTGAAGAAATGACCGGCGGAGCTGAACCGAGCAACGCTACTGTA 703
 DB 30084 CGGCGAATTTACTGAAAAAGATGACAGGCTGTAATGAACTTCAACGCTACTGTT 30025
 QY 704 GACGTGACATCCGCGCTATTCGTAATAATTCGATTCACCGCGATACCGCGAATC 763
 DB 30024 GATGTAAAGTTCGTGATTCGTAATAATTCGATTCACCGCGATACCGCGAATC 29965
 QY 764 ATGCCACCATTCACGCTGAAGGTTATCGCTTCTGCGGTGATGGAAGATTATGCGCT 823
 DB 29964 ATGCCACCATTCACGCTGAAGGTTATCGCTTCTGCGGTGATGGAAGATTATGCGCT 29905
 QY 824 TTACGACCGTCAAAAA 840
 DB 29904 CTATCGCTCTTAACA 29888
 RESULT 6
 ID ADJ38390 standard; DNA; 759 BP.
 AC ADJ38390;
 XX
 XX 06-MAY-2004 (first entry)
 DE
 XX Pantoea ananatis arca (aerobic respiration control) gene SeqID19.
 XX
 XX gamma-proteobacterium; Arca; aerobic respiration control;
 KM Arca protein function; chromosomal arca gene; fermentation; L-amino acid;
 KM L-lysine; L-glutamic acid; gene; ds.
 OS Pantoea ananatis.
 XX
 FH Key Location/Qualifiers

```

FT CDS 41..757
FT /*tag= a
FT /product= "Pantoea ananatis arca protein"
PD EPI382686-A1.
PD 21-JAN-2004.
XX
XX
XX
XX 11-JUL-2003; 2003EP-00015911.
XX
XX 12-JUL-2002; 2002JP-00203764.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Iehikawa Y, Imaizumi A, Matsui K, Kojima H;
XX WPI; 2004-192975/19.
XX P-PSDB; ADJ38391.
XX
XX New bacterium useful for producing a target substance (especially L-
XX lysine, L-glutamic acid or another L-amino acid) is modified so that the
XX Arca protein does not normally function.
XX
XX Example 4; SEQ ID NO 19; 38bp; English.
XX
XX This invention relates to a novel gamma-proteobacterium able to produce a
XX target substance such as an amino acid which is modified so that an Arca
XX (aerobic respiration control) protein does not normally function. The
XX disruption to Arca protein function is preferably due to disruption of a
XX chromosomal arca gene. The bacterium and method are useful in production of a
XX target substance by fermentation. The target substance is preferably an L-
XX -amino acid, especially L-lysine or L-glutamic acid. Production of the
XX target substance using the modified bacterium is more efficient. The
XX present sequence is that of the P ananatis gene which encodes the Arca
XX protein and which was used in the exemplification of the invention.
XX
XX Sequence 759 BP; 205 A; 192 C; 183 G; 179 T; 0 U; 0 Other;
SQ

```

```

Query Match 54.9%; Score 508.6; DB 12; Length 759;
Best Local Similarity 80.6%; Pred. No. 2.8e-158;
Matches 607; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

```

```

QY 69 TTTGATTGATGTTGGC--AATTAGGTAGCAAAATGAGACCCGCGCATTTCTTATGCT 126
DB 7 TTTCAATTATGTTGGCAAAATTAGTGTAAATGACGACCCGCGCATTTCTATGCT 66
QY 127 TGAAGACGAGTTGTGAACGCAACGTTGAAAGTATTTTGAAGCGGAAGCTATGA 186
DB 67 TGAAGACGAGTTGTGAACGCAACGTTGAAAGTATTTTGAAGCGGAAGCTATGA 186
QY 187 TGTTCGAAAGCGACGATGCGCGGAAATGATCAGATCTCTGTAATATGACATCAA 246
DB 127 CGGTACGAAAGCGACGATGCGCGGAAATGATCAGATCTCTGTAATATGACATCAA 186
QY 247 CCTGTGATCATGATATCAATCTGCGGTTAAAGCGTCTTCTGTTAGCGCGTAACT 306
DB 187 TCTGTATTAATGAGATCAATCTGCGGTTAAAGCGTCTTCTGTTAGCGCGTAACT 306
QY 307 GCGCGACGCGGAAATGTTGCGTTGATGTTCTGTAAGTGAAGCGGATGATGAT 246
DB 247 GCGTGAAGCGGAAATGTTGCGTTGATGTTCTGTAAGTGAAGCGGATGATGAT 306
QY 367 AATTCTGCGCTCGAAATGCGTGAAGTATCAATCAATCAACCGTTCAACCGCGTGA 426
DB 307 AATTCTGCGCTCGAAATGCGTGAAGTATCAATCAATCAACCGTTCAACCGCGTGA 426
QY 427 ACTGACGATTTGTGACGCAACTACTGCTCCGTAACCATGAATGCGGTACTGTACGCA 486
DB 367 ATTAACTATTGTCGACGTAACCTGCTGTGCGACCATGATTTGCTTTTACCATGA 426
QY 487 AGAAGCTGTAGCGTTGAAAGCTTACAGTTAATGTTGGGAATGACATCAACGCGG 546
DB 427 AGAAGCTGTAGCGTTGAAAGCTTACAGTTAATGTTGGGAATGACATCAACGCGG 486

```

```

QY 547 TTGCTTGATGCGCCCTGATGAGAGATGACAGTGAAGCTGCGCGCAGAGTATCCGCCAT 606
DB 487 CTACATCATCAATCCCAAGCGTGAAGTGAACAAATGCGCGCAGTGAATTCGATCCAT 546
QY 607 GCTTCACTTCTGTGAAACCCAGGCAAAATTAGTCCCGTGTGAACGTCTGAAGAAAT 666
DB 547 GTGCACTTCTGTGAAACCCAGGCAAAATTAGTCCCGTGTGAACGTCTGAAGAAAT 606
QY 667 GACCGCGCGTGAAGTGAACCCAGGCAAAATTAGTCCCGTGTGAACGTCTGAAGAAAT 726
DB 607 GACCGCGCGTGAAGTGAACCCAGGCAAAATTAGTCCCGTGTGAACGTCTGAAGAAAT 666
QY 727 TAAACATTTGATGATCAAGCGCGGATGAGCGGAAATGATGCGCAATTCAGCGTGAAG 786
DB 667 TAAACATTTGATGATCAAGCGCGGATGAGCGGAAATGATGCGCAATTCAGCGTGAAG 726
QY 787 TTAATGCTTTTGTGCGGTGATCTTGAAGATTATC 819
DB 727 TTAATGCTTTTGTGCGGTGATCTTGAAGATTATC 759

```

```

RESULT 7
ACF69844
ID ACF69844 standard; DNA; 717 BP.
AC ACF69844;
XX
XX 20-NOV-2003 (first entry)
XX
XX Photobhabdus luminescens nucleotide sequence #8311.
XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough; gene; ds.
XX
XX Photobhabdus luminescens.
XX
XX WO20029467-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002MO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;
XX Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photobhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 8311; 1205bp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photobhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and

```

CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which *P.*
CC *luminescens* is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated *P. luminescens* genes
XX
XX Sequence 717 BP; 213 A; 158 C; 158 G; 188 T; 0 U; 0 Other;
SQ
Query Match 53.2%; Score 493; DB 10; Length 717;
Best Local Similarity 80.5%; Pred. No. 4.2e-153;
Matches 577; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 101 ATGACAGCCCGGACATTTCTTATGTTGAAGACGATTTGTTAAACAGCAACAGCTGAA 160
DB 1 ATCCAAACCCCGACATTTTGTATTTGAAGACGAATTTCTCTCCCAATACCTTAA 60
QY 161 AGTATTTTGAAGCGGAGGCTATGATGTTTTCGAACGACAGATGGCCGGAATGCAT 220
DB 61 AGCATTTTGAAGCTGAGGAGGTACATAGTTTATGAAGCCACTGATGGTTCAGAAATGCAC 120
QY 221 CAGATCTCTCTGAATATGATCAATCAACCTGTGTATCATGATATCAATTCGCCGGTAA 280
DB 121 CATATTCTGTCAAAACAACGACATTAACCTGTGTATTAATTAACCTTCCAGGCAAA 180
QY 281 AACGGCTTCTGTAGCGCGTAACTGCGGACGAGCGCAATGTGGTTGATGTTCTCG 340
DB 181 AATGCTTGTACTTGTCCCGCAACTGCGAGACGAAACAAATTTGTTGATGTTCTCG 240
QY 341 ACTGCGCGTGAACAACGAGTGCATTAATTTCTGGCTCGAAATCGGTGAGTGA 400
DB 241 ACTGCGCGTGAATACGAAGTGAACAAATCTTGGCTGAAATCGGTGAGTGA 300
QY 401 ATCACCAAAACGTTCAACCGCGGTGAATCAATTCGTGACAGCACTCATGTCCT 460
DB 301 ATCACCAAAACATTTTATCGCGGTGAATGATCAATTCGCGCCGCAACCTGTTCTCG 360
QY 461 ACCATGATCTGGGTACTGTCGACGGAAGACGTGTAGGTTGAAGCTACATGATCAAT 520
DB 361 ACCATGATCTGTGATATGTCAGCGAGGAACGTGTAGGTTGAAGTATTAATTCAC 420
QY 521 GGTGGGAACGTGACATCAACAGCCGTGATGATGCGCCCTGATGCGGACGATCAAG 580
DB 421 GGTGGGAATTAATCAATCAATAGCCGTTCTGTATAGCCGCGCTGTGAGCCGTATAA 480
QY 581 CTGCGCGGACGAGGTTCCGCGCATGCTTCACTTCTGTGAACCCAGGCAAAATTCAG 640
DB 481 TTGCGCAGCAGAGGTTCCGCGCATGCTTCACTTCTGTGAATTCAGGCAAAATTCAG 540
QY 641 TCCCGGTGAACTGCTGAAGAAATGACCGCGCGTGAAGTGAACCGCAACGCTTACT 700
DB 541 ACGCGGACGATTTACTGAAAAAGATGACAGGTGTGAATTAACCTTCAACGCGTACT 600
QY 701 GTAGACGTGACGATCCGCGTATTCGTAAACATTTGGAATCTACGCGGATAGCGCGAA 760
DB 601 GTTGAATGAAGATTTGTGTATTCGTAAACATTTTGAATGACGCGCAATTAAGAGGAA 660
QY 761 ATCATGCGCAACATTCAGCGGTGAAGTATCGCTTCTGCGGTATCTGAAGATTA 817
DB 661 ATTATGCGCACATTCACGCGTGAAGGCTATCGTTCTGTGTGATTTGGAAGATTA 717

RESULT 8
ADP01748
ID ADP01748 standard; DNA; 726 BP.
XX
AC ADP01748;
XX
DT 12-FEB-2004 (first entry)
XX
DE Bacterial polynucleotide #2033.

XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW Immunosuppressant; gene; da.
XX
OS Proteus mirabilis.
XX
PN US6605709-B1.
XX
PN 12-AUG-2003.
XX
PD 05-APR-2000; 2000US-00543681.
PF
XX
XX 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL;
XX
XX WPI; 2003-895291/82.
DR P-PSDB; ADP05920.
XX
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 2033; 870bp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunizing an individual against *P. mirabilis*, a
CC method for evaluating a compound for the ability to bind a *P. mirabilis*
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polynucleotide of the invention.
XX
XX Sequence 726 BP; 224 A; 137 C; 149 G; 216 T; 0 U; 0 Other;
SQ
Query Match 48.6%; Score 450.8; DB 10; Length 726;
Best Local Similarity 76.3%; Pred. No. 5e-139;
Matches 554; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 92 GTAGCAACATGCAAGCCCGCAACATTTCTTATGTTGAAGACGATGTTGAACGCAAC 151
DB 1 GTAGCAAAATATGCAAAACCCGCAATTTGATTTGAAGATGAAGTACTCTGTAAT 60
QY 152 ACGTTGAAGAAATTTTTCGAACGGAAGGCTATGATGTTTGAAGCGGACATGCGCGG 211
DB 61 ACCCTGAAAGAAATTTTTCGAACGGAAGGCTATGATGTTTGAAGCGGACATGCGCGG 120
QY 212 GAATGATCATGATCCTCTCTGAATATGATCAACCTGTGATCATGATATCAATCTG 271
DB 121 GATGATGCAATATTTCTGTGCGACATGATATCAATCTGTGATATGATATTAATCTT 180
QY 272 CCGGTTAAGAACGCTTCTGTGTTAGCGCGTGAACCTGCGGACGAGCGAATGTTGCTTG 331
DB 181 CCGTGAAGAAATGCTTCTTATTTAGCCGCTGAATTAACGGAACAGGTAGTGTGATTA 240
QY 332 ATGTTCTGATCGGCGGTGACCAACGATGCTAATAATCTCGGCTCGAAATCGGTGA 391
DB 241 ATGTTCTTAACAGGTGTGATATGAAAGTTAATAATCTTGAAGCTTGAATTTGTTGTC 300
QY 392 GATGATCATGACCAACCGTTCAACCGCGTGAACGATTCGTCACGCAACCTA 451
DB 301 GATGATTAATCATCAATTAACATTTTAACTCGTGAATTAACATCGTGTCTGTACTTA 360
QY 452 CTGTCCCGTACATGATCTGGGTACTGTACGGAAGAACTGTGATGCTTGAAGCTTAC 511

Db 361 TTGTACGCGATATGATTTAGCGAATGGCAGAGAAGCGTGTATTGATTGAAAGCTAT 420
QY 512 AAGTTCATGTTGGGAACTGGACATCAACAGCCGCTGCTGATGCGCCGATGGGCGAG 571
Db 421 AAATTTAATGTTGGAGCTGATATTAATAGTCCCTCTTATTAAGCCCTACAGGTGAA 480
QY 572 CAGTCAACAGTCCCGCGCAGGAGTTCGCGCCGATGCTTCACTTCTGTAACCCAGGC 631
Db 481 CAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
QY 632 AAAATTCAGTCCCGCGCAGGAGTTCGCGCCGATGCTTCACTTCTGTAACCCAGGC 691
Db 541 AAAATTCAGTCCCGCGCAGGAGTTCGCGCCGATGCTTCACTTCTGTAACCCAGGC 600
QY 692 GACGCTACTGTAAGCTGACGATCCGCGGATGCTTCACTTCTGTAACCCAGGC 751
Db 601 GATGCTACTGTAAGCTGACGATCCGCGGATGCTTCACTTCTGTAACCCAGGC 660
QY 752 ACGCGGAAATCATCGGACCATTCACGCTGTAAGCTTATGCTTCTGCGGTGATCTGAA 811
Db 661 ACACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
QY 812 GATTAA 817
Db 721 GAGTGA 726

RESULT 9

AAS92939 ID AAS92939 standard; cDNA; 1048 BP.

AC AAS92939;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #28743.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG28752.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnosis, forensics, gene mapping, identification of mutations

PT biotechnology.

PS Claim 1; SEQ ID NO 28743; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SO Sequence 1048 BP; 297 A; 265 C; 283 G; 203 T; 0 U; 0 Other;
Query Match 48.3%; Score 447.4; DB 5; Length 1048;
Best Local Similarity 99.8%; Pred. No. 8.3e-138;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 312 AGCAGCGCAATGTTGCTGATGTTCTGACTGCGCGTGACAGCAAGTGTGATAAATTC 371
Db 573 AACAGCGCAATGTTGCTGATGTTCTGACTGCGCGTGACAGCAAGTGTGATAAATTC 632
QY 372 TCGGCTCGAAATCGGTGACAGTACATCACCAACCGTTCAACCGCGTGAATGCA 431
Db 633 TCGGCTCGAAATCGGTGACAGTACATCACCAACCGTTCAACCGCGTGAATGCA 692
QY 432 CGATTGTCGACGCAACCTACTGTCCTGATCATGTAATCTGGTACTGTGACGCAAGAC 491
Db 693 CGATTGTCGACGCAACCTACTGTCCTGATCATGTAATCTGGTACTGTGACGCAAGAC 752
QY 492 GTCCGAGGCTGGAAGCAATCAATGTTCAATGTTGGAACTGAGCATCAAGCCGTTCT 551
Db 753 GTCCGAGGCTGGAAGCAATCAATGTTCAATGTTGGAACTGAGCATCAAGCCGTTCT 812
QY 552 TGATCGGCTGATGAGGAGAGTACAAAGCTGCGCGCAGCGAGTTCGCGGCAATGCTTC 611
Db 813 TGATCGGCTGATGAGGAGAGTACAAAGCTGCGCGCAGCGAGTTCGCGGCAATGCTTC 872
QY 612 ACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGTAAGAAATGACCG 671
Db 873 ACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGTAAGAAATGACCG 932
QY 672 GCCGTGAGCTGGAACCGCAGCGGATCTGTAAGCAAGTCCGCGGATCTGTAAGCAAG 731
Db 933 GCCGTGAGCTGGAACCGCAGCGGATCTGTAAGCAAGTCCGCGGATCTGTAAGCAAG 992
QY 732 ATTTGGAATCTACGCGCGGATACCGCGGAA 760
Db 993 ATTTGGAATCTACGCGCGGATACCGCGGAA 1021

RESULT 10

AAT42063_09/c Continuation (10 of 19) of AAT42063 from base 900001 (Haemophilus influenzae complete ge

WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063
WP Fragment Name Begin End
WP AAT42063_00 1 110000
WP AAT42063_01 100001 210000
WP AAT42063_02 200001 310000
WP AAT42063_03 300001 410000
WP AAT42063_04 400001 510000
WP AAT42063_05 500001 610000
WP AAT42063_06 600001 710000
WP AAT42063_07 700001 810000
WP AAT42063_08 800001 910000
WP AAT42063_09 900001 1010000
WP AAT42063_10 1000001 1110000
WP AAT42063_11 1100001 1210000
WP AAT42063_12 1200001 1310000
WP AAT42063_13 1300001 1410000
WP AAT42063_14 1400001 1510000

WP AAT42063_15 1500001 1610000
WP AAT42063_16 1600001 1710000
WP AAT42063_17 1700001 1810000
WP AAT42063_18 1800001 1830121

Query Match 34.5%; Score 319.4; DB 2; Length 110000;

Best Local Similarity 65.8%; Pred. No. 5,4e-94;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTAGTAGCAACATGACAGACCCGCACTTCTATGCTGAAGACAGTTGGTAAC 144
DB 36640 AGATAGAGATGAGAAATGACTCTCCAAAATTCCTGTTGAGATGATAATTTGTAC 36581
QY 145 AGCAACACGTTGAAAAGATTTCGAAGCGGAAGCTATGATGTTTCGACGACAGA 204
DB 36580 TCGAATATACGCTTAAGGAGATTTTGAAGCGGAAGATGATGTTTGAACAGAAA 36521
QY 205 TGGCGGGAATGATCATCATCTCTCTGTAATATGACATCACTGCTGATCATGATAT 264
DB 36520 TGGTGTGAATGATCATCATATATGGAATCATATATATTTGGTGTGATGATAT 36461
QY 265 CATCTGCGGGGTAAGACGGCTCTGTTAGCGCGGAAGTGGAGAGCGGAGATG 324
DB 36460 TAATTTACAGGCAAAACGGCTTATTTGSCAAGAACTCGTGAAGATTAAGCTT 36401
QY 325 TCGGTGATGTTCTCTGATGCGCGGTGACAGAGAGTGAATAATTTCCGCGCTCGAAAT 384
DB 36400 ACTCTTATTTTAACTGCTGAGATATGAAAGTGAATAATTTTGGTTGGAAT 36341
QY 385 CGGTGAGATGATCATCATCAACCAACCGTTCAACCGCGGTGAAGTGAAGTGTGACG 444
DB 36340 TGGTCCGATGACTATTTGCTAAGCTTTTAACTGAGAACTGATCTATCCGCGCAGC 36281
QY 445 CAACCTACTGTCGCGTACCAAGATCTGAGAGAGAAAGTGTGAGCGTTGA 504
DB 36280 TAATCTATTGATCGTGAATG---CCGATCAAGAAAAGAAATATCATTTGTGAGAG 36224
QY 505 AAGCTCAAGTTCAATGTTGGAGATGACATCAACAGCGCTGCTGATCGCGCTCGA 564
DB 36223 ATTCTATCGTTTAAATGTTGAAATTAAGCTTAATTAAGCAAGTTTATTAACACAG 36164
QY 565 TGGCGACAGTACAGCTGCGCGGACGAGATTCGCGCGATGCTTCACTTCTGTAAGA 624
DB 36163 AGGACACAGATTCAACTTCTGCGAGTGAATTCGTGCAATGTTCAATTCGTGAAA 36104
QY 625 CCAGGCAAAATCACTCCCGTCTAACTGAGAGAAATTAACCGCGCGTGAAGTGA 684
DB 36103 TCCAGGAAATGCAAAACGGCGAAGATTTGAAAAAATTAACCGAGAGAGTTAA 36044
QY 685 ACCGACGACCGTACTGATGACGTGACATCCGCGTATTCGTAACATTTGCAATCTAC 744
DB 36043 ACTCAGAGATCGATCGATGATGTCACAAATTCGACGATACAGAAACATTTGAAGATC 35984
QY 745 GCGGATACCGCGGAATATCGCCACCATTCACGCGTGAAGTTATGCTTGCAGTGA 804
DB 35983 TCCCAATACCTCCAAATATCATTTGACTATACATGAGAAAGCTATGTTTGTGAGA 35924
QY 805 TCTGGAAGA 813
DB 35923 TAATGAATA 35915

RESULT 11
AAS53409

ID AAS53409 standard; DNA; 711 BP.

AC AAS53409;

DT 13-FEB-2002 (first entry)

DE Haemophilus influenzae DNA for cellular proliferation protein #191.

KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;

KW antibacterial; drug design.
OS Haemophilus influenzae.
XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

XX 23-MAY-2000; 2000US-0206848P.

XX 26-MAY-2000; 2000US-0207727P.

XX 23-OCT-2000; 2000US-0242578P.

XX 27-NOV-2000; 2000US-0253625P.

XX 22-DEC-2000; 2000US-0257931P.

XX 16-FEB-2001; 2001US-0269308P.

XX (ELITRA PHARM INC.

XX Haseelbeck R, Ohlsen KL, Zyskind JM, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX P-PSDB; A035550.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids.

XX Claim 27; SEQ ID NO 7046; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes,
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX CC nucleic acids which are required for cell proliferation in a wide variety
XX CC of organisms. The present sequence encodes an essential prokaryotic
XX CC cellular proliferation protein. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 711 BP; 246 A; 113 C; 143 G; 209 T; 0 U; 0 Other;

Query Match 34.1%; Score 316.2; DB 4; Length 711;
Best Local Similarity 66.2%; Pred. No. 3,6e-94;
Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 101 ATGCAAGACCCGCACTTCTTATCGTTGAAGCAGTGTGTAACGCAACGCTTGAA 160

DB 1 ATGACTACTCCAAAATTTCTGTTGAAGATGAATTTGCTCGAATATGCTTAA 60

QY 161 AGTATTTGACGCGAAGCTATGATGTTTTCGAAGCAGATGCGCGGAATGCAT 220

DB 61 GGGATTTTGAAGCGGAAGATGATGTTTTCGAAGCAGAAATGTTGAAATGCAT 120

QY 221 CAGATCTCTGATATGACATCAACCTGGTATGATGATATCAATCTGCCGGTAA 280

DB 121 CATATTTGCAAAATCATATATTAATTTGGTGTGATGATATTAATTTACGAGCA 180

QY 281 AACGCTCTCTGTTAGCGGTGAACGCGGACGAGCGCAAGTGGCGTTGATGTTCC 340

DB 181 AACGCTTATTTATTTGCAAGAGAACTCGTGAAGATTAAGCTTACTTATTTT 240

QY 341 ACTGCGGTGACAGAACTGATTAATTTCTGCGCTCGAATCGGTGACATGACTAC 400

Db 241 ACTGTCGAGTATGAAAGTGAATTAATTTAGTTGGAATTTGTGCGATGACTAT 300
Qy 401 ATCAACCAACCGTTCAACCGCGTGAACGATTCGTGACGCAACCTACTGTCCTG 460
Db 301 TTGACTAAGCTTTTAACCTTAAGAACTGACTATCCGGCAGTAACTATGATCGT 360
Qy 461 ACCATGAATTTGGGTACTGTCAACGGAAGACGTGAGCGTTGAAGCTACAAAGTTCAT 520
Db 361 GCAATG---CCGCATCAAGAAAAAGAAATTAATTTGTCGAGAAATTTCTATGTTTAT 417
Qy 521 GATTGGAACTGGAATCAACAGCCCTGTTGATGCGCCCTGATGCGGACGATCAAG 580
Db 418 GATTGAAATTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 477
Qy 581 CTGCGCGGACGAGTTCGCGCGCATGCTTCACTTCTGTAAGAAACCGGCAAAATTCAG 640
Db 478 CTTCCTCGAGTGAATTTGTCGCAATTTCTGTAAGAAATTCGAAAAATTCGAA 537
Qy 641 TCCCGTCTGAACCTGCTGAAGAAATGACCGCGCGTGAAGCTGAAACCGCAAGCGTACT 700
Db 538 ACGCGGAGAAATTTGCTGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 597
Qy 701 GTAGACGTGACGATCCGCGTATGCTTAAATTTGCAATCTACGCGGATACCGCGGAA 760
Db 598 GTAGATGTCAATTCGAGATTCGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 657
Qy 761 ATCATGCGCAACATTCAGCGTGAAGTATGCTTCTGCGGTGATCTGGAAGA 813
Db 658 ATCATTAATGACTAATGATGAGAAAGGCTATGTTTGTGGAATATGATA 710

RESULT 12
ACA34287
ID ACA34287 standard; DNA; 711 BP.
AC ACA34287;
XX
XX
XX 19-JUN-2003 (first entry)
DT
XX
DE Prokaryotic essential gene #15944.
DE
XX
KW Antisense; ds: prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Haemophilus influenzae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU30417.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 22157; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs; or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
SQ
Sequence 711 BP; 246 A; 113 C; 143 G; 209 T; 0 U; 0 Other;

Query Match 34.1%; Score 316.2; DB 8; Length 711;
Best Local Similarity 66.2%; Pred. No. 3, 6e-94;
Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

Qy 101 ATGCAGACCCCGACATTTCTTATCGTTGAAGACGATGCTGAACGCAACGTTGAA 160
Db 1 ATGACTATCTCAAAAATTCGTTGTAAGTGAATGTGCTACGAAATGCGTTAA 60
Qy 161 AGTATTTTCAGACCGAAGGCTATGATGTTTTCGAAGCAGATGCGCGGAATGCAT 220
Db 61 GGGATTTTTCAGACCGAAGGATATGATGTTTTCGAAGCAGAAATGTTGAATATGCAT 120
Qy 221 CAGATCTCTCTGAATATGACATCACTGTGATCATGATCATCATGCGCGGATAG 280
Db 121 CATATATGCAAAATCATATATTAATTTGTTGATGATGATATTAATTACAGCAAA 180
Qy 281 AACGCTCTCTGTAGCGGTGAACGCGGCGAGCGAAGTGTGCTTATGTTCTG 340
Db 181 AACGCTTATTAATGCGAAGAACTCGTGAAGATTAATTAATTAATTTTAA 240
Qy 341 ACTGCGCGTGAACGAAAGTGAATTAATTCGCGCTGAAATCGGTGACATGATAC 400
Db 241 ACTGTCGAGATTAAGAGTGAATTAATTTAGGTTGGAATTTGTCGCGATGACTAT 300
Qy 401 ATCAACCAACCGTTCAACCGCGTGAACGATTCGTGACGCAACCTACTGTCCTG 460
Db 301 TTGACTAAGCTTTTAACCTTAAGAACTGACTATCCGGCAGTAACTATGATCGT 360
Qy 461 ACCATGAATTTGGGTACTGTCAACGGAAGACGTGAGCGTTGAAGCTACAAAGTTCAT 520
Db 361 GCAATG---CCGCATCAAGAAAAAGAAATTAATTTGTCGAGAAATTTCTATGTTTAT 417
Qy 521 GATTGGAACTGGAATCAACAGCCCTGTTGATGCGCCCTGATGCGGACGATCAAG 580
Db 418 GATTGAAATTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 477
Qy 581 CTGCGCGGACGAGTTCGCGCGCATGCTTCACTTCTGTAAGAAACCGGCAAAATTCAG 640
Db 478 CTTCCTCGAGTGAATTTGTCGCAATTTCTGTAAGAAATTCGAAAAATTCGAA 537

QY 641 TCCCGTCTGAACCTGCTGAAGAAATGACCGCCGTGAGTGAAACCGCAGACCGTACT 700
DB 538 ACCCGCGAAGAAATGCTGAAAATAATGACCGAGAGATTAATAACTCAGATCGTACC 597
QY 701 GTGAGAGTGAACGATCCGCCGTATTCGTAACATTTTCATCTACGCCGAGTACGCCGGA 760
DB 598 GTGAGATGTCACAAATTCGACCTATCAGAAAACATTTTGAGATCATCCCAATCTCCAAAT 657
QY 761 ATCATGCGCCACATTCGACGCTGATCGCTTCGCGGTATCTGAGAGA 813
DB 658 ATCATTAATGACTATACATGAGAGAGCTATCGTTTGTGAGATATTGAATA 710

RESULT 13
AAS82422
ID AAS82422 standard; cDNA; 1110 BP.
XX AAS82422;
AC AAS82422;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18226.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG18235.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 18226; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed Specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1110 BP; 256 A; 291 C; 296 G; 267 T; 0 U; 0 Other;
Query Match 25.1%; Score 232.8; DB 5; Length 1110;
Best Local Similarity 98.8%; Pred. No. 2.9e-66;
Matches 245; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 544 CCGTTCGTTGATCGGCCCTGATGCGAGCAGTA-CAAGCTGCCCGCAGGAGTTCCGCG 602
DB 761 CCGTTGTTGATATGGCCCTGATGGCGAGGTAACAAGCTGCCCGCAGGAGTTCCGCG 820
QY 603 CCAATGCTTCACTTCTGTGAACACCCAGCAAAATTGATGCCGTCTGAACTGCTGAAGA 662
DB 821 CCATGCTTCACTTCTGTGAACACCCAGCAAAATTGATGCCGTCTGAACTGCTGAAGA 880
QY 663 AATGACCGCGCGGTGAGTGAACCCGACGACCGTACTGAGACGTGAGATCGCGCGTA 722
DB 881 AATGACCGCGCGGTGAGTGAACCCGACGACCGTACTGAGACGTGAGATCGCGCGTA 940
QY 723 TTGTAACATTTTGAATCTACCGCGGATACGCCGGAATTCATCGCCACCATTCACGGTG 782
DB 941 TTGTAACATTTTGAATCTACCGCGGATACGCCGGAATTCATCGCCACCATTCACGGTG 1000
QY 783 AAGCTTAT 790
DB 1001 AAGTTCAT 1008

RESULT 14
ABQ90117
ID ABQ90117 standard; DNA; 729 BP.
XX
XX ABQ90117;
AC
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #102 for DNA array.
XX
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
PN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002MO-NO000019.
XX
PR 12-JAN-2001; 2001NO-00000235.
PR 12-JAN-2001; 2001NO-00000239.
XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillhaug JR, Lossius I, Bisen UB, Frazer CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557818/59.
XX
XX Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes.
XX
XX
PS Claim 19; Page 99; 678bp; English.
XX
XX The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the

CC invention
XX Sequence 729 BP; 136 A; 241 C; 222 G; 130 T; 0 U; 0 Other;
SQ
Query Match 14.8%; Score 137.4; DB 6; Length 729;
Best Local Similarity 49.8%; Pred. No. 1.4e-34;
Matches 348; Conservative 0; Mismatches 351; Indels 0; Gaps 0;
QY 104 CAGACCCCGGCACTTCTTATGTTGAGAGCAGTGTGTAACACCGCAACAGCTTAAAGT 163
DB 10 CCGACACACAGTAATCTGATGTGATGATGACGACCTCCAGATTCGGAACTCTCTCGCGCA 69
QY 164 ATTTTCAGAGCGGAGGATGATGATGTTTTCAGACGACAGATGCGCGGAAATGCATCAG 223
DB 70 TACCTTCGCGCTTCGCGATGAGAGCCGCTGCTTCATGATGCGCGGATGAAAG 129
QY 224 ATCTCTCTGAATATGACATCACTGATGATGATGATGATGATGATGATGATGATGATGAT 283
DB 130 GCGCTGACGGAATGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 189
QY 284 GGTCTTCTGTTTACGCGCTGAACTGCGGACGAGCGAATGTTGCTGATGTTCTGACT 343
DB 190 GGCCTGACCTGTCGCGCTGCTCGCTGCGGAGAGCAGCACTCCGTCATCATCTGACG 249
QY 344 GCGCGTGAACAAGAGTCGATTAATCTCGGCTCGAATTCGGTGAATGATGATGATGATGATGAT 403
DB 250 GCGAGGCGGATCCATGACACCGGCTGCTGCGCTGGAATCTCGCGCGGACGATGATGATGAT 309
QY 404 ACCAACCCTTCAACCCGCTGAACTGACGATGCTGACGACGACCACTGATGCTCGGTAAC 463
DB 310 GCCAAGCGTTTCAAGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369
QY 464 ATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523
DB 370 CCGGACGAGGCGCAACCGGAAACCCCGAATCTCCGACATCCGCTGGAATTCGAAGG 429
QY 524 TGGGAACTGACATCAACAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 583
DB 430 TGGGCGGTGACCGGGGTGCTCCGCACTGACCTCCGCAAGGCAATGCTGCTCTCTC 489
QY 584 CCGGCGAGCGATGTCGCGGCACTGCTTCACTTCTGTAACCCAGCAAAATTCAGTCC 643
DB 490 TCCATATGCCAGATGTCGCGCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 549
QY 644 CGTCTGAAGCTGCTGAAGAAATGACCGGCTCGTGAAGTGAACCGCAACGCTGATGTA 703
DB 550 CCGGATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
QY 704 GACGTGACGATCCGCGTATGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGAT 763
DB 610 GACCTGCTGATGCTCCGCGTGGCGGCAAGATGACCCCAAGATCCCAATCCCAATC 669
QY 764 ATGCGCAACATTCAGCTGAAGTTATTCCTTCTGCGGT 802
DB 670 ATCAAGACCATCGCGCGAGGTTATTTCTTACGCGCT 708
RESULT 15
ABD17813
ID ABD17813 standard; DNA; 837 BP.
XX
AC ABD17813;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #16417.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US551795-B1.

XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
XX
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, DeLoughery C, Bush D;
XX
DR WPI; 2003-615309/58.
XX
DR P-PDSB; ABO84242.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX
XX pathological conditions resulting from bacterial infection.
PS Disclosure; SEQ ID NO 16417; 455bp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for recombinant
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD1397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
SQ Sequence 837 BP; 140 A; 303 C; 273 G; 121 T; 0 U; 0 Other;
Query Match 13.6%; Score 126; DB 11; Length 837;
Best Local Similarity 52.9%; Pred. No. 9.7e-31;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;
QY 243 TCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302
DB 263 TCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 322
QY 303 AACTGCGGAGCAGCGCAATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 362
DB 323 AGCTGCGGATGCTCGGAGGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 382
QY 363 ATTAATTTCTGCGCTGGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 422
DB 383 ATGCAATGCTGCGCTGGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442
QY 423 GTGAATGACGATGCTGCAACGCACTGATCCGCTACATGAAATCTGATGATGATGATGATGATGAT 482
DB 443 GCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
QY 483 GCGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 542
DB 500 GCGCGGCGCCCGCGCGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 559
QY 543 GCGCTGAT 602
DB 560 GCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 619
QY 603 CCAATGCTTCACTTCTGTAACCCAGGCAAAATTCAGTCCGCTGATGATGATGATGATGATGATGAT 662
DB 620 TGCTCGGCGCTTCTGCGGACAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 679
QY 663 AAATGACCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722

Db	680	AGATCCGCAACCGGAGTGGCTGCCCCAGCATCGCTTCATCGAGTGTGATCGGCC	739
Qy	723	TTGGTAAACATTTGAACTACGCGGATACCGGAAATCATGCGCACCATTCACGGTG	782
Db	740	TGGTGGCAAGCTGCGGAGAGACCGGCCGAACTGATCATCACATCCACGGCG	799
Qy	783	AAGTTATCGCTTCTGCG	800
Db	800	CCGGTACCTGTTCACCG	817

Search completed: January 22, 2005, 12:42:09
 Job time : 556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 22, 2005, 12:15:23 ; Search time 105 Seconds
(without alignments)
6275.250 Million cell updates/sec

Title: US-10-613-990A-31

Perfect score: 927
Sequence: 1 gtcacgtcaccgcgcacatcatg.....gtacctctctactggtctt 927

Scoring table: IDENTITY NUC
Gapox 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	577.2	62.3	726	4 US-09-489-039A-1812	Sequence 1812, Ap
2	450.8	48.6	726	4 US-09-543-681A-2033	Sequence 2033, Ap
3	319.4	34.5	1830121	4 US-09-557-884-1	Sequence 1, Appli
4	319.4	34.5	1830121	4 US-09-643-990A-1	Sequence 1, Appli
5	319.4	34.5	1830121	4 US-10-329-960-1	Sequence 1, Appli
6	126	13.6	837	4 US-09-252-991A-16417	Sequence 16417, A
7	126	13.6	2430	4 US-09-252-991A-16171	Sequence 16171, A
8	126	13.6	3675	4 US-09-252-991A-16309	Sequence 16309, A
9	120.6	13.0	990	4 US-09-252-991A-11517	Sequence 11517, A
10	120.6	13.0	2166	4 US-09-252-991A-11722	Sequence 11722, A
11	113.8	12.3	2004	4 US-09-252-991A-11865	Sequence 11865, A
12	107.4	11.6	1208	4 US-09-418-980-1	Sequence 1, Appli
13	98.6	10.6	729	4 US-09-634-238-133	Sequence 133, App
14	97.2	10.5	762	4 US-09-252-991A-6307	Sequence 6307, App
15	96.2	10.4	705	3 US-09-134-001C-226	Sequence 226, App
16	94.6	10.2	702	4 US-09-710-279-479	Sequence 479, App
17	94.6	10.2	3389	4 US-09-710-279-3914	Sequence 3914, Ap
18	92.6	10.0	782	4 US-09-252-991A-10543	Sequence 10543, A
19	92.6	10.0	804	4 US-09-252-991A-10323	Sequence 10323, A
20	91.4	9.9	702	4 US-09-082-077-2	Sequence 2, Appli
21	91.4	9.9	3731	4 US-09-082-077-1	Sequence 1, Appli
22	91.4	9.9	9425	4 US-08-956-171E-87	Sequence 87, Appli
23	91.4	9.9	9425	4 US-08-781-986A-87	Sequence 87, Appli
24	87.6	9.4	783	4 US-09-489-039A-3722	Sequence 3722, Ap
25	87.4	9.4	945	4 US-09-252-991A-3092	Sequence 3092, Ap
26	87.4	9.4	978	4 US-09-252-991A-2878	Sequence 2878, Ap
27	87.4	9.4	2523	4 US-09-252-991A-2990	Sequence 2990, Ap

28	85.4	9.2	825	4 US-09-489-039A-1321	Sequence 1321, Ap
29	85.2	9.2	738	4 US-09-252-991A-14821	Sequence 14821, A
30	85.2	9.2	1341	4 US-09-252-991A-14543	Sequence 14543, A
31	84.8	9.1	738	4 US-09-489-039A-1129	Sequence 1129, Ap
32	84.4	9.1	9909	4 US-08-961-527-12	Sequence 12, Appli
33	83.2	9.0	1060	1 US-08-850-118-1	Sequence 1, Appli
34	83.2	9.0	1060	2 US-09-008-253-1	Sequence 1, Appli
35	83.2	9.0	1060	3 US-09-093-335-1	Sequence 1, Appli
36	83	9.0	774	4 US-09-252-991A-5924	Sequence 5924, Ap
37	82.2	8.9	881	4 US-09-252-991A-5118	Sequence 5118, Ap
38	82.2	8.9	981	4 US-09-252-991A-5003	Sequence 5003, Ap
39	80.8	8.7	705	4 US-09-583-110-1923	Sequence 1923, Ap
40	79.6	8.6	9306	3 US-09-453-702B-231	Sequence 231, App
41	79.2	8.5	606	4 US-09-489-039A-1368	Sequence 1368, Ap
42	79.2	8.5	702	4 US-09-489-039A-5026	Sequence 5026, Ap
43	78.2	8.4	1717	4 US-09-634-238-6	Sequence 6, Appli
44	78	8.4	985	4 US-09-489-039A-2595	Sequence 2595, Ap
45	77.8	8.4	2064	4 US-09-252-991A-15079	Sequence 15079, A

ALIGNMENTS

RESULT 1									
US-09-489-039A-1812									
; Sequence 1812, Application US/09489039A									
; Patent No. 6610836									
; GENERAL INFORMATION:									
; APPLICANT: GARY BRETON et. al									
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA									
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS									
; CURRENT APPLICATION NUMBER: US/09/489,039A									
; PRIOR FILING DATE: 2000-01-27									
; PRIOR APPLICATION NUMBER: US 60/117,747									
; NUMBER OF SEQ ID NOS: 14342									
; SEQ ID NO 1812									
; LENGTH: 726									
; TYPE: DNA									
; ORGANISM: Klebsiella pneumoniae									
US-09-489-039A-1812									
Query Match 62.3%; Score 577.2; DB 4; Length 726;									
Best Local Similarity 87.2%; Pred. No. 1,1e-197;									
Matches 633; Conservative 0; Mismatches 93; Indels 0; Gaps 0;									
QY	92	GTAGCAACATGAGACCCCGACATTTCTTATGTTGAAGACGAGTTGGTAACGCCAAC	151						
DB	1	GTAGCAACATGAGACCCCGACATTTCTTATGTTGAAGACGAGTTGGTAACGCCAAC	60						
QY	152	ACGTTAAAGATTTTTCGAAGCGGAGCTATGATGTTTTCGAACGACAGATGGCGG	211						
DB	61	ACGTTAAAGATTTTTCGAAGCGGAGCTATGATGTTTTCGAACGACAGATGGCGG	120						
QY	212	GAATGATCATGATCTCTCTGTAATGATCATCAACCTGATCATGATATCAATCTG	271						
DB	121	GAATGATCATGATCTCTCTGTAATGATCATCAACCTGATCATGATATCAATCTG	180						
QY	272	CCGGGTAAAGACGTTCTTGTAGCGGTGAACCTGCGGACGAGCGAATGTTGCTTG	331						
DB	181	CCGGGTAAAGATGTTCTTGTAGCGGTGAACCTGCGGACGAGCGAATGTTGCTTG	240						
QY	332	ATGTTCTGATCGTCCGTCGACCAAGAGTCATTAATTTCTGGCTTCAAAATCGGTGA	391						
DB	241	ATGTTCTGATCGTCCGTCGACCAAGAGTCATTAATTTCTGGCTTCAAAATCGGTGA	300						
QY	392	GATGATCATCATCAACCGTTCAACCGCGGATGATGATGATGATGATGATGATGATG	451						
DB	301	GATGATCATCATCAACCGTTCAACCGCGGATGATGATGATGATGATGATGATGATG	360						
QY	452	CTGTCCGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	511						

Db	361	CTCTCCCGACATGAAACCTCGGTACCGTAGCGAAGAGCGTCGACGCTGGAAAAGCTAC	420
Oy	512	AAGTTCAATGTTGGGAACCTGACATCAACAGCCGTTGTTGATCGAGCCCTGATGGCAG	571
Db	421	AAGTTCAACGCTGGGAACTCGATATCAACACGCCCTTCCCTGGTTAGCGCCCAACGGCGMA	480
Oy	572	CAGTACAAAGCTGCGCGGACGAGAGTTCCGCGCATGTTCACTTCTGTGAAAAACCCAGGC	631
Db	481	CAGTACAAAGCTGCGCGGACGAGATTCGCGCATGTTCACTTCTGTGAAAAACCCAGGC	540
Oy	632	AAATTCAGTCCCGGTCGTACCTGCTGAAAGAAAATGACCCGCGCTGAGCTGAAAACGCGAC	691
Db	541	AAATTCAGTCTGTGTGTCGAGCTGCTGAAAGAAAATGACCCGCGCTGAGCTGAAAACGCGCAT	600
Oy	692	GACCGTACTGTGAGCTGACGATCCGCGCTATTGCTAAACATTTCGAATCTACGCGCGAT	751
Db	601	GACCGTACGTTGACCTGACCATTTGTGGGATGCTAAACACTTCGAATCCACTCCGAT	660
Oy	752	ACGCGGAAATATCGCCACCATTAAGGTAAAGTTACGCTCTGCGGTGATCTGGAA	811
Db	661	ACCGCGAAATATCTGCTACCATTAACGTAAAGGCTACCGTTCTGTGGCGATCTGCAG	720
Oy	812	GATTA 817	
Db	721	GAATTA 726	

RESULT 2

```

US-09-543-681A-2033
/ Sequence 2033, Application US/09543681A
/ Patent No. 6605709
/ GENERAL INFORMATION:
/ APPLICANT: GARY BRETON
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
/ TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 2709.1002-001
/ CURRENT APPLICATION NUMBER: US/09/543,681A
/ CURRENT FILING DATE: 2000-04-05
/ PRIOR APPLICATION NUMBER: US 60/128,706
/ PRIOR FILING DATE: 1999-04-09
/ NUMBER OF SEQ ID NOS: 8344
/ SEQ ID NO 2033
/ LENGTH: 726
/ TYPE: DNA
/ ORGANISM: Proteus mirabilis
/ US-09-543-681A-2033

```

Query Match	48.6%	Score 450.8;	DB 4;	Length 726;
Best Local Similarity	76.3%	Pred. No. 5e-152;		
Matches 554; Conservative	0;	Mismatches 172;	Indels 0;	Gaps 0

QY 92 GTAGCAAAACATGCAGACCCCGCACTTCTTATCGTTGAAGACAGATTGGTAAACAGCAAC 151

Db 1 GTAGCAAAATATCAAAACCCCGCACTTCTGATGTGTTGAAGATGAAGTAGTACTCGTAAT 60

QY 152 ACGTTGAAAAGATATTTGGAACGGAGGCCTATGATGTTTTCGAAAGCGACAGATGGCGG 211

Db 61 ACCCTGAAAAGCATTTGGAAGCTAAGGGTATATCGACAGAAAGCCATGATGGCAAC 120

QY 212 GAAATGCATCAGATCTCTCTTGAATATGACATCAACCTGGATCATGATCAATCAATCTG 271

Db 121 GAGATGCATTAATATCTCTCGACCATGATATCAATCTGTCATTTATGATTAATTAATCTT 180

QY 272 CCGGGTAAGAACGGCTCTCTGTATACGGGTAAACCTGGCCGACAGCGCAATGTTGCGTTG 331

Db 181 CCTGGTAAAAATGGCTCTTCTTAATACCCGTGAATTAACGTGAACAGTAAAGTTGGCATTA 240

QY 332 ATGTTCTGACTGCGCCGTGACACAGAACTGCATTAATAATTCGCGCTCGAATCGATCA 391

Db 241 ATGTTCTTAAACGGCTCGATTAATGAATGATTAATAATCTTAAGCCTTGAAATTTGGTACC 300

QY 392 GATGACTATCATCAACAAACGGTTAACCCTGGTGAACGATGATTCGTGCACGCAACCTTA 451

Db	30.	GATGATTAACATCACTAAACCATTTAATCTCGTAATTAACTAATCGTGCTCTGAACCTTA	360
QY	452	CTGTCCCGTACCATGAATCTGGGACTGTCAACGAGAACGTCGTAGCTGTGAAGCTAC	511
Db	361	TTGTCAACGACATATGATATTTAGCGAATGGCACAGAAAGCGTGCTTATGTTGAAGCTAT	420
QY	512	AAGTCAATAGTGTGGGAAGTGGACATCAACAGCCGTCGTTGATCGGCCCTGATGGCAG	571
Db	421	AAATTTAAATGTGGAGCTGAATATTAATATGTCGCTCTCTTTATACCCCTACAGGTGAA	480
QY	572	CAGTACAAAGCTGCGCGACGAGATTCGCGCCCAATGCTTCACTTCTGTGAAAAACGAGC	631
Db	481	CAGTATTAATATCACTCGTAATGAGATTTGTGCGAATGTTAATTTCTGCGAAAAACGAGA	540
QY	632	AAAAATTCAGTCCCGTGCTGAACTGCTGAAGAAAAATGAACCGGCGGTGAGTGAACCGCAC	691
Db	541	AAAATCCAAATCTGTGCAAGAAATTACTGAAAAAATGACGGGTGTGAATTAATAACTCAT	600
QY	692	GACCGTACTGTGACGTAGCATCGCGCGTATTCGTGAACAATTGCAATTTACGCCGAT	751
Db	601	GATTCGATCGTAGAGCTTACCATTCGTCGTGTAATTCGTAACACTTTAAATCAACCCCTGAT	660
QY	752	ACGCCGAAATCATGCCACCAATTCACGGTGAAGTTATGCTTCTGCGGTGATCTGAA	811
Db	661	ACACTGTGATTAATGCACTATCAATGAGGAAGTTATGTTTCTGTGTGATATTTAGAC	720
QY	812	GATTAA 817	
Db	721	GAGTGA 726	

RESULT 3

US-09-557-884-1/c / Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.4e-102;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

85 AATTAGTAGCAACATGAGACCCGACATCTTATCGTTGAGACGATGTTGTTAC 144
DB 936640 AGATATAGATGAGAAATGATCTCTCAAAAATTTCTGTTGAGATGAAATTTGCAC 936581
QY 145 ACCGACACGTTGAAAAGTATTTTTCGAGCGGAAAGCTATGATGTTTTCGACGACAGA 204
DB 936580 TCGAATATAGCTTAAAGGGATTTTTCGAGCGGAAAGATGATGTTTTCGACGAGAAA 936521
QY 205 TGGCGGGAATGATCATGATCTCTCTGTAATGATCATGACCTGATGATGATAT 264
DB 936520 TGGTGTGAAATGATCATATATATGGAATCATATATATATATATATATATATATATAT 936461
QY 265 CAATCTGCGGGTAGAAGCGCTCTCTGTTAGCGCGTGAACCTGCGGACGAGCGAATGT 324
DB 936460 TAATTTACAGGCAAAAAGCGCTTATATATGGAAGGAACTCGGTGAAGAAATTAAGCTT 936401
QY 325 TGGCTGATGTTCTCTGACCTGCGCTGACAAAGATGATTAATTTCTCGGCTCGAAT 384
DB 936400 ACCTCTAATTTTAACTGCTGAGATATGAAAGTGAATAATTTAGTTTGAAT 936341
QY 385 CGGTGAGATGATCATATCAACAAACCGTTCAACCGCGGATGATGATGATGATGATGATGAT 444
DB 936340 TGGTGCAGATGATCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 936281
QY 445 CAACCTACTGTCCTGATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
DB 936280 TAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936224
QY 505 AAGCTCAAGTTCAATGTTGGGAACTGACATCAACGCGCTTCTGTTGATCGGCTCTGA 564
DB 936223 ATCTATCGTTTAAAGTTGGAATTAAGACCTAAATAGCAAGTTTAAATTAACACAGA 936164
QY 565 TGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 624
DB 936163 AGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936104
QY 625 CCGAGCAAAATTCAGTCCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 684
DB 936103 TCCAGGAAATTTCCAAACGCGGCAAGATGCTGAAAAAATGACCGGACGAGAGTTAAA 936044
QY 685 ACCGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744
DB 936043 ACCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 935984
QY 745 GCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804
DB 935983 TCCCAATATCTCAAAATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 935924
QY 805 TCTGAGAA 813
DB 935923 TATTGAATA 935915

RESULT 4

US-09-643-990A-1/c
Sequence 1. Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann

Mark D. Adams
Owen White
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.4e-102;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

85 AATTAGTAGCAACATGAGACCCGACATCTTATCGTTGAGACGATGTTGTTAC 144
DB 936640 AGATATAGATGAGAAATGATCTCTCAAAAATTTCTGTTGAGATGAAATTTGCAC 936581
QY 145 ACCGACACGTTGAAAAGTATTTTTCGAGCGGAAAGCTATGATGTTTTCGACGACAGA 204
DB 936580 TCGAATATAGCTTAAAGGGATTTTTCGAGCGGAAAGATGATGTTTTCGACGAGAAA 936521
QY 205 TGGCGGGAATGATCATGATCTCTCTGTAATGATCATGACCTGATGATGATAT 264
DB 936520 TGGTGTGAAATGATCATATATATGGAATCATATATATATATATATATATATATATAT 936461
QY 265 CAATCTGCGGGTAGAAGCGCTCTCTGTTAGCGCGTGAACCTGCGGACGAGCGAATGT 324
DB 936460 TAATTTACAGGCAAAAAGCGCTTATATATGGAAGGAACTCGGTGAAGAAATTAAGCTT 936401
QY 325 TGGCTGATGTTCTCTGACCTGCGCTGACAAAGATGATTAATTTCTCGGCTCGAAT 384
DB 936400 ACCTCTAATTTTAACTGCTGAGATATGAAAGTGAATAATTTAGTTTGAAT 936341
QY 385 CGGTGAGATGATCATATCAACAAACCGTTCAACCGCGGATGATGATGATGATGATGATGATGAT 444
DB 936340 TGGTGCAGATGATCATATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 936281
QY 445 CAACCTACTGTCCTGATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
DB 936280 TAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936224
QY 505 AAGCTCAAGTTCAATGTTGGGAACTGACATCAACGCGCTTCTGTTGATCGGCTCTGA 564
DB 936223 ATCTATCGTTTAAAGTTGGAATTAAGACCTAAATAGCAAGTTTAAATTAACACAGA 936164

QY 565 TGGCGAGCATGACAGCTGCGCGCAGAGTTCCGCGCATCTTCTCTGTAAA 624
Db 936163 AGGACAGAAATTCAACTTCTCGCATGAAATTCGCAATTTCTGTAAAA 936104
QY 625 CCCAGGCAAAATTCAGTCCCTGCTGTAAGTCTGAAAGAAATGACCGCGCTGAGCTGAA 684
Db 936103 TCCAGGAAATTCGAAACGCGGAGAAATTCGTGAAAAAATGACCGGAGAGATTAA 936044
QY 685 ACCGCGACGACGCTAGCTGACGATCGCGCGGATTCGTAACATTTGGAATCTAC 744
Db 936043 ACCTCAGAGATCGTACCGTAGATGCACAATTCGACGTAACGAAACATTTGGAATCA 935984
QY 745 GCGGATAGCGCGGAAATCATCCCAACATTCACGCTGAGGTTATCGCTTCTGCGGTGA 804
Db 935983 TCCCAATCTCCAAATATCATTAATGACTATACATGAGAGGCTATCGTTTGTGAGAA 935924
QY 805 TCTGGAAGA 813
Db 935923 TATTGAATA 935915

RESULT 5

US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Patent No. 6742927
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; Patent No. 6742927
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIORITY FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)

OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248) .. (107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136) .. (117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750) .. (119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924) .. (119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038) .. (120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344) .. (121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336) .. (122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167) .. (122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340) .. (131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360) .. (131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910) .. (139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398) .. (140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750) .. (142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058) .. (145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171) .. (145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942) .. (145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197) .. (147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841) .. (150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```

```

NAME/KEY: misc_feature
LOCATION: (152500) .. (152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4.4e-102;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

QY 85 AATTAGTAGCAAAATGAGACCCCGACATCTTATCGTTGGAAGAGAGTTGGTAAC 144
DB 936640 AGTATAGAGTAGAAATGACTCTCCAAAATTCGTGTGTGAAGATGAAATGTGCAC 936581
QY 145 ACCCAACAGCTTGAAGAAATTTTCGAAGCGAAGGCTATGATGTTTTCGAACGACAGA 204
DB 936580 TCGAAATACGCTTAAAGGATTTTTCGAAGCGAAGATATGATGTGTTGAACAGAGAA 936521
QY 205 TGGCGCGAAATGATCAATCCTCTCTGAAATATGACATCAACCTGATGATCATATAT 264
DB 936520 TGGTGTGAATGATCATATATATGCAAAATCATATATATATGTTGTGATGATAT 936461
QY 265 CAATCTGCGGGTGAAGAGCGTTCTGTTAGGCGGTGAACCTGCGGAGAGCGAAATG 324
DB 936460 TAATTTACCAAGCAAAACGGCTTATTTATTTGCAAGAACTCCGTGAAGAAATTAAGCTT 936401
QY 325 TGGCTGATGTTCTCTGATGCGCGGTGACAAAGATGATTAATTCGCGCTCGAAAT 384
DB 936400 ACCTCTATTTTATTTTAACTGCTGATGATTAATGAAAGTGAATTAATTTAGGTTTGAAT 936341
QY 385 CGGTGAGATGACTATACATCAACCAACCGTTCAACCGCGGTGAACCTGATGCTGACG 444
DB 936340 TGGTGCAGATGACTATTTGATCTAAGCCTTTTAACTGAAGAACTGATCTGCGGACG 936281
QY 445 CAACCTACGTCCTCCGATCAATGATCTGGTACTGTCAGAGAAAGACGTGACGTGA 504
DB 936280 TAATCTATGATGATGCAATG--CCGATCAAGAAAGAAATATTTGTCGAGA 936224
QY 505 AAGCTACAAGTTCAATGATGTTGGAACTGACATCAACAGCGCTTGTGATCGGCCCTGA 564
DB 936223 ATTCTATCGTTTAAATGTTGAAATTAAGACCTTAATAGCACAGTTTAATTACACAGA 936164
QY 565 TGGCGAGCATACACCTGCGCGGACGAGAGTTCCGCGCATGCTTCACTTCTGTGAAA 624
DB 936163 AGGACAAAGATTCAACTCTCTGAGTGAATTTGTCGATGATTCATTTCTGTGAAA 936104
QY 625 CCCAGCAAAATTCAGTCCCGTGTCTGAATCTGTAAGAAATGACCGGCGGTGAGCTGAA 684
DB 936103 TCCAGGAAATTCGCAACCGCGCAAGAAATTCGTAAGAAATGACCGGACGAGAGTTAA 936044
QY 685 ACCGACGACCGTACGTAGAGTGAAGTCCGCGTATTCGTAACATTTGATGATCTAC 744
DB 936043 ACCTCAGATCGTACCGTGAAGTGCACAAATTCAGCATATCAGAAACATTTTGAAGATCA 935984
QY 745 GCCGATACCGCGGAATATCATGCCACATTCAGCGTGAAGTTATCGCTTCTGCGGTGA 804
DB 935983 TCCCATATCTCCAAATATCATATATGATCATATGAGAGAGGTATCGTTTGTGGAAG 935924
QY 805 TCTGGAAGA 813
DB 935923 TATTAATA 935915

RESULT 6
US-09-252-991A-16417
; Sequence 16417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16417
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16417

Query Match      13.6%; Score 126; DB 4; Length 837;
Best Local Similarity 52.9%; Pred. No. 1,1e-34;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY 243 TCAACCTGGTATCATATGATATCAATCTGCGGGTAAAGAACGCTCTTCTGTAGCCGCTG 302
DB 263 TCGAGCTGTGTCTGTGATCGATCCGCTGCGGCAAGACGCGCTGACCCCTGACCCGCG 322
QY 303 AACTGCGCGACGACGCGGATGTTGCGTTGATGTTCTGACTGCGCCGCGTGAACAAGTGC 362
DB 323 AGCTGGGGGTGGCTCCGAGGTGGGATCATCTGTATCACCGGGCCGACAGAGATCG 382
QY 363 ATAAATTTCTGGCTCTGAAATCGGTGACATGATCAATCAACAAACCGTTCAACCGCG 422
DB 383 ATGCACTGTGTGCGCTGAGTGGCGGCGCGACGATTACGTATCAAGCGCTGACCCGCG 442
QY 423 GTGACATGACGATTTGTCGACGCAACCTATCTCCCTGACATGATCTGGGTACTGTCA 482
DB 443 GCGAATCTGTGTGCGGCGC--GAAATATGTATCCGCGGAGCGCCATGCGCCAGGCCA 499
QY 483 GCGAAGACGTCTGTAGCGTTGAAAGCTACAAAGTTCAATGTTGGAAGTGAACATCAACA 542
DB 500 GCGCGCGCGCGCGCGGAGGCGCTCAGGCAAGTTCCGCGCATGTGCTGTGAGCGCCGAC 559
QY 543 GCGCTTCTGTATCGGCGCTGATGCGAGCAGTCAAGCTCCGCGCAGAGAGTTCCGCG 602
DB 560 GGGCGCGCGCTGATGACACGCGGCGACGAAACCTCTGCTCAACCTAGCGAATTCAT 619
QY 603 CCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGTGAATCTGTAAAGA 662
DB 620 TGTGCGCGCTTCTGTGCGGACAGCGGCGATACCTTGAGCGCGCAGTGTGATGAGCC 679
QY 663 AAATGACGCGCGCTGTGAGTAAACCGCACGACCGTACTGTAGACGTGACGATCCGCGTA 722
DB 680 AGATCGCAACCGCGGAGTGGCTGCCAGGATGCTCATGACGTGTGTGCGCGCGC 739
QY 723 TTGCTAAACATTTCAATCTAGCGCGGATACGCGGAAATCATCGCCACATTCACGCTG 782
DB 740 TGGGTGCAAGCTGCGCGACGACCCGCGCAACCGCACTGATCATCAATCCACGCGCG 799
QY 783 AAGTTATCGCTTCTGCG 800
DB 800 CCGGCTACTGTTCAACG 817

RESULT 7
US-09-252-991A-16171/c
; Sequence 16171, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 16171
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16171

Query Match      13.6%; Score 126; DB 4; Length 2430;
Best Local Similarity 52.9%; Pred. No. 2,4e-34;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY 243 TCAACCTGGTATCATATGATATCAATCTGCGGGTAAAGAACGCTCTTCTGTAGCCGCTG 302
DB 1406 TCGAGCTGTGTCTGTGATCGATCCGCTGCGGCAAGACGCGCTGACCCCTGACCCGCG 1347
QY 303 AACTGCGCGACGACGCGGATGTTGCGTTGATGTTCTGACTGCGCCGCGTGAACAAGTGC 362
DB 1346 AGCTGGGGGTGGCTCCGAGGTGGGATCATCTGATCAACCGGGCGCAACGAGATCG 1287
QY 363 ATAAATTTCTGGCTCTGAAATCGGTGACATGATCAATCAACAAACCGTTCAACCGCG 422
DB 1286 ATGCACTGTGTGCGCTGAGTGGCGCGCGACGATTAAGTATCAAGCGCTGAACCGCG 1227
QY 423 GTGAACTGACGATTTGTCGACGCAACCTATGTTCCGTAACATGAATCTGGGTACTGTCA 482
DB 1226 GCGAATGATGTGTGCGGCGC--GAAATATGTATCCGCGGAGTGGCGCATGCGCCAGGCCA 1170
QY 483 GCGAAGACGTGTGATGCGTTGAAACCTACAAAGTTCAATGTTGGGAATGACATCAACA 542
DB 1169 GCGCGCGCGCGCGCGGAGCGCTTGAAGCAATTTGGGACATGCTGTGTCGACCGCGAC 1110
QY 543 GCGCTTCTGTATCGGCGCTGTATGCGAGCAGTACAACTGCGCGCGAGCGAGTTCCGCG 602
DB 1109 GCGCGCGCTGTATTCACACGCGGCAACGAACCTGTCTACCCATGCGCAATTCAT 1050
QY 603 CCATCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGTGCTGAATCTGTAAAGA 662
DB 1049 TGTGCGCGCTTCTGTGCGCAACAGCGGCGATACCTGAGCGCGACGATGATGAGCC 990
QY 663 AAATGACGCGCGCTGTGAGTAAACCGCACGACCGTACTGTGACGTGACATCCGCGTA 722
DB 989 AGATCGCAACCGCGAGTGGCTGCCAGCGATGCTCATGACGTGTGTGCGCGCGC 930
QY 723 TTGCTAAACATTTCAATCTAGCGCGGATACGCGGAAATCATCGCCACATTCACGCTG 782
DB 929 TGGGTGCAAGCTGCGGACGACCGCGCAACCGCACTGATCATCAATCCACGCGCG 870
QY 783 AAGTTATCGCTTCTGCG 800
DB 869 CCGGCTACTGTTCAACG 852

RESULT 8
US-09-252-991A-16309
; Sequence 16309, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16309
; LENGTH: 3675
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16309
```